

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
2	2	1	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
3	3	2	1	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
4	4	3	2	1	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
5	5	4	3	2	1	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80																				

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
5	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80																				

Monsanto Company
Brian Stierwalt
c/o Gail Wuellner
Mail Code E2NA
800 N. Lindbergh Blvd.
St. Louis, MO 63167
(636) 737-7286

Carl J. Schwedler
Reg. No. 36,924

Enclosure

NUCLEIC ACID SEQUENCES TO PROTEINS INVOLVED IN TOCOPHEROL SYNTHESIS

5

INTRODUCTION

TECHNICAL FIELD

10 The present invention is directed to nucleic acid and amino acid sequences and
constructs, and methods related thereto.

BACKGROUND

15 Isoprenoids are ubiquitous compounds found in all living organisms. Plants synthesize a
diverse array of greater than 22,000 isoprenoids (Connolly and Hill (1992) *Dictionary of
Terpenoids*, Chapman and Hall, New York, NY). In plants, isoprenoids play essential roles in
particular cell functions such as production of sterols, contributing to eukaryotic membrane
architecture, acyclic polyprenoids found in the side chain of ubiquinone and plastoquinone,
growth regulators like abscisic acid, gibberellins, brassinosteroids or the photosynthetic pigments
chlorophylls and carotenoids. Although the physiological role of other plant isoprenoids is less
20 evident, like that of the vast array of secondary metabolites, some are known to play key roles
mediating the adaptative responses to different environmental challenges. In spite of the
remarkable diversity of structure and function, all isoprenoids originate from a single metabolic
precursor, isopentenyl diphosphate (IPP) (Wright, (1961) *Annu. Rev. Biochem.* 20:525-548; and
Spurgeon and Porter, (1981) in Biosynthesis of Isoprenoid Compounds., Porter and Spurgeon eds
25 (John Wiley, New York) Vol. 1, pp1-46).

A number of unique and interconnected biochemical pathways derived from the
isoprenoid pathway leading to secondary metabolites, including tocopherols, exist in chloroplasts
of higher plants. Tocopherols not only perform vital functions in plants, but are also important
from mammalian nutritional perspectives. In plastids, tocopherols account for up to 40% of the
30 total quinone pool.

Tocopherols and tocotrienols (unsaturated tocopherol derivatives) are well known antioxidants, and play an important role in protecting cells from free radical damage, and in the prevention of many diseases, including cardiac disease, cancer, cataracts, retinopathy, Alzheimer's disease, and neurodegeneration, and have been shown to have beneficial effects on symptoms of arthritis, and in anti-aging. Vitamin E is used in chicken feed for improving the shelf life, appearance, flavor, and oxidative stability of meat, and to transfer tocopherols from feed to eggs. Vitamin E has been shown to be essential for normal reproduction, improves overall performance, and enhances immunocompetence in livestock animals. Vitamin E supplement in animal feed also imparts oxidative stability to milk products.

The demand for natural tocopherols as supplements has been steadily growing at a rate of 10-20% for the past three years. At present, the demand exceeds the supply for natural tocopherols, which are known to be more biopotent than racemic mixtures of synthetically produced tocopherols. Naturally occurring tocopherols are all *d*-stereoisomers, whereas synthetic α -tocopherol is a mixture of eight *d,l*- α -tocopherol isomers, only one of which (12.5%) is identical to the natural *d*- α -tocopherol. Natural *d*- α -tocopherol has the highest vitamin E activity (1.49 IU/mg) when compared to other natural tocopherols or tocotrienols. The synthetic α -tocopherol has a vitamin E activity of 1.1 IU/mg. In 1995, the worldwide market for raw refined tocopherols was \$1020 million; synthetic materials comprised 85-88% of the market, the remaining 12-15% being natural materials. The best sources of natural tocopherols and tocotrienols are vegetable oils and grain products. Currently, most of the natural Vitamin E is produced from γ -tocopherol derived from soy oil processing, which is subsequently converted to α -tocopherol by chemical modification (α -tocopherol exhibits the greatest biological activity).

Methods of enhancing the levels of tocopherols and tocotrienols in plants, especially levels of the more desirable compounds that can be used directly, without chemical modification, would be useful to the art as such molecules exhibit better functionality and bioavailability.

In addition, methods for the increased production of other isoprenoid derived compounds in a host plant cell is desirable. Furthermore, methods for the production of particular isoprenoid compounds in a host plant cell is also needed.

SUMMARY OF THE INVENTION

The present invention is directed to sequences to proteins involved in tocopherol synthesis. The polynucleotides and polypeptides of the present invention include those derived
5 from prokaryotic and eukaryotic sources.

Thus, one aspect of the present invention relates to prenyltransferase, and in particular to isolated polynucleotide sequences encoding prenyltransferase proteins and polypeptides related thereto. In particular, isolated nucleic acid sequences encoding prenyltransferase proteins from bacterial and plant sources are provided.

10 In another aspect, the present invention provides isolated polynucleotide sequences encoding tocopherol cyclase, and polypeptides related thereto. In particular, isolated nucleic acid sequences encoding tocopherol cyclase proteins from bacterial and plant sources are provided.

Another aspect of the present invention relates to oligonucleotides which include partial or complete prenyltransferase or tocopherol cyclase encoding sequences.

15 It is also an aspect of the present invention to provide recombinant DNA constructs which can be used for transcription or transcription and translation (expression) of prenyltransferase or tocopherol cyclase. In particular, constructs are provided which are capable of transcription or transcription and translation in host cells.

20 In another aspect of the present invention, methods are provided for production of prenyltransferase or tocopherol cyclase in a host cell or progeny thereof. In particular, host cells are transformed or transfected with a DNA construct which can be used for transcription or transcription and translation of prenyltransferase or tocopherol cyclase. The recombinant cells which contain prenyltransferase or tocopherol cyclase are also part of the present invention.

25 In a further aspect, the present invention relates to methods of using polynucleotide and polypeptide sequences to modify the tocopherol content of host cells, particularly in host plant cells. Plant cells having such a modified tocopherol content are also contemplated herein. Methods and cells in which both prenyltransferase and tocopherol cyclase are expressed in a host cell are also part of the present invention.

30 The modified plants, seeds and oils obtained by the expression of the prenyltransferase or tocopherol cyclase are also considered part of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 provides an amino acid sequence alignment between ATPT2, ATPT3, ATPT4,
5 ATPT8, and ATPT12 are performed using ClustalW.

Figure 2 provides a schematic picture of the expression construct pCGN10800.

Figure 3 provides a schematic picture of the expression construct pCGN10801.

Figure 4 provides a schematic picture of the expression construct pCGN10803.

Figure 5 provides a schematic picture of the construct pCGN10806.

10 Figure 6 provides a schematic picture of the construct pCGN10807.

Figure 7 provides a schematic picture of the construct pCGN10808.

Figure 8 provides a schematic picture of the expression construct pCGN10809.

Figure 9 provides a schematic picture of the expression construct pCGN10810.

Figure 10 provides a schematic picture of the expression construct pCGN10811.

5 Figure 11 provides a schematic picture of the expression construct pCGN10812.

Figure 12 provides a schematic picture of the expression construct pCGN10813.

Figure 13 provides a schematic picture of the expression construct pCGN10814.

Figure 14 provides a schematic picture of the expression construct pCGN10815.

Figure 15 provides a schematic picture of the expression construct pCGN10816.

20 Figure 16 provides a schematic picture of the expression construct pCGN10817.

Figure 17 provides a schematic picture of the expression construct pCGN10819.

Figure 18 provides a schematic picture of the expression construct pCGN10824.

Figure 19 provides a schematic picture of the expression construct pCGN10825.

Figure 20 provides a schematic picture of the expression construct pCGN10826.

25 Figure 21 provides an amino acid sequence alignment using ClustalW between the
• *Synechocystis* prenyltransferase sequences.

Figure 22 provides an amino acid sequence of the ATPT2, ATPT3, ATPT4, ATPT8, and
ATPT12 protein sequences from *Arabidopsis* and the slr1736, slr0926, slr1899, slr0056, and the
slr1518 amino acid sequences from *Synechocystis*.

Figure 23 provides the results of the enzymatic assay from preparations of wild type *Synechocystis* strain 6803, and *Synechocystis* slr1736 knockout.

Figure 24 provides bar graphs of HPLC data obtained from seed extracts of transgenic *Arabidopsis* containing pCGN10822, which provides of the expression of the ATPT2 sequence, in the sense orientation, from the napin promoter. Provided are graphs for alpha, gamma, and delta tocopherols, as well as total tocopherol for 22 transformed lines, as well as a nontransformed (wildtype) control.

Figure 25 provides a bar graph of HPLC analysis of seed extracts from *Arabidopsis* plants transformed with pCGN10803 (35S-ATPT2, in the antisense orientation), pCGN10822 (line 1625, napin ATPT2 in the sense orientation), pCGN10809 (line 1627, 35S-ATPT3 in the sense orientation), a nontransformed (wt) control, and an empty vector transformed control.

Figure 26 shows total tocopherol levels measured in T# *Arabidopsis* seed of line.

Figure 27 shows total tocopherol levels measured in T# *Arabidopsis* seed of line.

Figure 28 shows total tocopherol levels measured in developing canola seed of line 10822-1.

Figure 29: shows results of phytyl prenyltransferase activity assay using *Synechocystis* wild type and slr1737 knockout mutant membrane preparations.

Figure 30 is the chromatograph from an HPLC analysis of *Synechocystis* extracts.

Figure 31 is a sequence alignment of the *Arabidopsis* homologue with the sequence of the public database.

Figure 32 shows the results of hydropathic analysis of slr1737

Figure 33 shows the results of hydropathic analysis of the *Arabidopsis* homologue of slr1737.

Figure 34 shows the catalytic mechanism of various cyclase enzymes

Figure 35 is a sequence alignment of slr1737, slr1737 *Arabidopsis* homologue and the *Arabidopsis* chalcone isomerase.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides, *inter alia*, compositions and methods for altering (for example, increasing and decreasing) the tocopherol levels and/or modulating their ratios in host cells. In particular, the present invention provides polynucleotides, polypeptides, and methods of use thereof for the modulation of tocopherol content in host plant cells.

5 The biosynthesis of α -tocopherol in higher plants involves condensation of homogentisic acid and phytylpyrophosphate to form 2-methyl-6 phytylbenzoquinol that can, by cyclization and subsequent methylations (Fiedler et al., 1982, *Planta*, 155: 511-515, Soll et al., 1980, *Arch. Biochem. Biophys.* 204: 544-550, Marshall et al., 1985 *Phytochem.*, 24: 1705-1711, all of which are herein incorporated by reference in their entirety), form various tocopherols.

10 The *Arabidopsis pds2* mutant identified and characterized by Norris *et al.* (1995), is deficient in tocopherol and plastiquinone-9 accumulation. Further genetic and biochemical analysis suggested that the protein encoded by *PDS2* may be responsible for the prenylation of homogentisic acid. The *PDS2* locus identified by Norris *et al.* (1995) has been hypothesized to possibly encode the tocopherol phytyl-prenyltransferase, as the *pds2* mutant fails to accumulate tocopherols.

15 Norris *et al.* (1995) determined that in *Arabidopsis pds2* lies at the top of chromosome 3, approximately 7 centimorgans above long hypocotyl2, based on the genetic map. ATPT2 is located on chromosome 2 between 36 and 41 centimorgans, lying on BAC F19F24, indicating that ATPT2 does not correspond to *PDS2*. Thus, it is an aspect of the present invention to provide novel polynucleotides and polypeptides involved in the prenylation of homogentisic acid. This reaction may be a rate limiting step in tocopherol biosynthesis, and this gene has yet to be isolated.

20 U.S. Patent No. 5,432,069 describes the partial purification and characterization of tocopherol cyclase from *Chlorella protothecoides*, *Dunaliella salina* and wheat. The cyclase described as being glycine rich, water soluble and with a predicted MW of 48-50kDa. However, only limited peptide fragment sequences were available.

25 In one aspect, the present invention provides polynucleotide and polypeptide sequences involved in the prenylation of straight chain and aromatic compounds. Straight chain prenyltransferases as used herein comprises sequences which encode proteins involved in the prenylation of straight chain compounds, including, but not limited to, geranyl geranyl

30

pyrophosphate and farnesyl pyrophosphate. Aromatic prenyltransferases, as used herein, comprises sequences which encode proteins involved in the prenylation of aromatic compounds, including, but not limited to, menaquinone, ubiquinone, chlorophyll, and homogentisic acid. The prenyltransferase of the present invention preferably prenylates homogentisic acid.

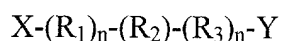
5 In another aspect, the invention provides polynucleotide and polypeptide sequences to tocopherol cyclization enzymes. The 2,3-dimethyl-5-phytylplastoquinol cyclase (tocopherol cyclase) is responsible for the cyclization of 2,3-dimethyl-5-phytylplastoquinol to tocopherol.

Isolated Polynucleotides, Proteins, and Polypeptides

10 A first aspect of the present invention relates to isolated prenyltransferase polynucleotides. Another aspect of the present invention relates to isolated tocopherol cyclase polynucleotides. The polynucleotide sequences of the present invention include isolated polynucleotides that encode the polypeptides of the invention having a deduced amino acid sequence selected from the group of sequences set forth in the Sequence Listing and to other polynucleotide sequences closely related to such sequences and variants thereof.

5 The invention provides a polynucleotide sequence identical over its entire length to each coding sequence as set forth in the Sequence Listing. The invention also provides the coding sequence for the mature polypeptide or a fragment thereof, as well as the coding sequence for the mature polypeptide or a fragment thereof in a reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, pro-, or prepro- protein sequence. The polynucleotide can also include non-coding sequences, including for example, but not limited to, non-coding 5' and 3' sequences, such as the transcribed, untranslated sequences, termination signals, ribosome binding sites, sequences that stabilize mRNA, introns, polyadenylation signals, and additional coding sequence that encodes additional amino acids. For example, a marker sequence can be included to facilitate the purification of the fused polypeptide. Polynucleotides of the present invention also include polynucleotides comprising a structural gene and the naturally associated sequences that control gene expression.

20 The invention also includes polynucleotides of the formula:



wherein, at the 5' end, X is hydrogen, and at the 3' end, Y is hydrogen or a metal, R₁ and R₃ are any nucleic acid residue, n is an integer between 1 and 3000, preferably between 1 and 1000 and R₂ is a nucleic acid sequence of the invention, particularly a nucleic acid sequence selected from the group set forth in the Sequence Listing and preferably those of SEQ ID NOs: 1, 3, 5, 7, 8, 10, 11, 13-16, 18, 23, 29, 36, and 38. In the formula, R₂ is oriented so that its 5' end residue is at the left, bound to R₁, and its 3' end residue is at the right, bound to R₃. Any stretch of nucleic acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer.

The invention also relates to variants of the polynucleotides described herein that encode for variants of the polypeptides of the invention. Variants that are fragments of the polynucleotides of the invention can be used to synthesize full-length polynucleotides of the invention. Preferred embodiments are polynucleotides encoding polypeptide variants wherein 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues of a polypeptide sequence of the invention are substituted, added or deleted, in any combination. Particularly preferred are substitutions, additions, and deletions that are silent such that they do not alter the properties or activities of the polynucleotide or polypeptide.

Further preferred embodiments of the invention that are at least 50%, 60%, or 70% identical over their entire length to a polynucleotide encoding a polypeptide of the invention, and polynucleotides that are complementary to such polynucleotides. More preferable are polynucleotides that comprise a region that is at least 80% identical over its entire length to a polynucleotide encoding a polypeptide of the invention and polynucleotides that are complementary thereto. In this regard, polynucleotides at least 90% identical over their entire length are particularly preferred, those at least 95% identical are especially preferred. Further, those with at least 97% identity are highly preferred and those with at least 98% and 99% identity are particularly highly preferred, with those at least 99% being the most highly preferred.

Preferred embodiments are polynucleotides that encode polypeptides that retain substantially the same biological function or activity as the mature polypeptides encoded by the polynucleotides set forth in the Sequence Listing.

The invention further relates to polynucleotides that hybridize to the above-described sequences. In particular, the invention relates to polynucleotides that hybridize under stringent

conditions to the above-described polynucleotides. As used herein, the terms “stringent conditions” and “stringent hybridization conditions” mean that hybridization will generally occur if there is at least 95% and preferably at least 97% identity between the sequences. An example of stringent hybridization conditions is overnight incubation at 42°C in a solution comprising 50% formamide, 5x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt’s solution, 10% dextran sulfate, and 20 micrograms/milliliter denatured, sheared salmon sperm DNA, followed by washing the hybridization support in 0.1x SSC at approximately 65°C. Other hybridization and wash conditions are well known and are exemplified in Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, NY (1989), particularly Chapter 11.

The invention also provides a polynucleotide consisting essentially of a polynucleotide sequence obtainable by screening an appropriate library containing the complete gene for a polynucleotide sequence set forth in the Sequence Listing under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence or a fragment thereof; and isolating said polynucleotide sequence. Fragments useful for obtaining such a polynucleotide include, for example, probes and primers as described herein.

As discussed herein regarding polynucleotide assays of the invention, for example, polynucleotides of the invention can be used as a hybridization probe for RNA, cDNA, or genomic DNA to isolate full length cDNAs or genomic clones encoding a polypeptide and to isolate cDNA or genomic clones of other genes that have a high sequence similarity to a polynucleotide set forth in the Sequence Listing. Such probes will generally comprise at least 15 bases. Preferably such probes will have at least 30 bases and can have at least 50 bases. Particularly preferred probes will have between 30 bases and 50 bases, inclusive.

The coding region of each gene that comprises or is comprised by a polynucleotide sequence set forth in the Sequence Listing may be isolated by screening using a DNA sequence provided in the Sequence Listing to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the invention is then used to screen a library of cDNA, genomic DNA or mRNA to identify members of the library which hybridize to the probe. For example, synthetic oligonucleotides are prepared which correspond to the prenyltransferase or tocopherol cyclase EST sequences. The oligonucleotides are used as

primers in polymerase chain reaction (PCR) techniques to obtain 5' and 3' terminal sequence of prenyltransferase or tocopherol cyclase genes. Alternatively, where oligonucleotides of low degeneracy can be prepared from particular prenyltransferase or tocopherol cyclase peptides, such probes may be used directly to screen gene libraries for prenyltransferase or tocopherol cyclase gene sequences. In particular, screening of cDNA libraries in phage vectors is useful in such methods due to lower levels of background hybridization.

Typically, a prenyltransferase or tocopherol cyclase sequence obtainable from the use of nucleic acid probes will show 60-70% sequence identity between the target prenyltransferase or tocopherol cyclase sequence and the encoding sequence used as a probe. However, lengthy sequences with as little as 50-60% sequence identity may also be obtained. The nucleic acid probes may be a lengthy fragment of the nucleic acid sequence, or may also be a shorter, oligonucleotide probe. When longer nucleic acid fragments are employed as probes (greater than about 100 bp), one may screen at lower stringencies in order to obtain sequences from the target sample which have 20-50% deviation (i.e., 50-80% sequence homology) from the sequences used as probe. Oligonucleotide probes can be considerably shorter than the entire nucleic acid sequence encoding an prenyltransferase or tocopherol cyclase enzyme, but should be at least about 10, preferably at least about 15, and more preferably at least about 20 nucleotides. A higher degree of sequence identity is desired when shorter regions are used as opposed to longer regions. It may thus be desirable to identify regions of highly conserved amino acid sequence to design oligonucleotide probes for detecting and recovering other related prenyltransferase or tocopherol cyclase genes. Shorter probes are often particularly useful for polymerase chain reactions (PCR), especially when highly conserved sequences can be identified. (*See, Gould, et al., PNAS USA (1989) 86:1934-1938.*).

Another aspect of the present invention relates to prenyltransferase or tocopherol cyclase polypeptides. Such polypeptides include isolated polypeptides set forth in the Sequence Listing, as well as polypeptides and fragments thereof, particularly those polypeptides which exhibit prenyltransferase or tocopherol cyclase activity and also those polypeptides which have at least 50%, 60% or 70% identity, preferably at least 80% identity, more preferably at least 90% identity, and most preferably at least 95% identity to a polypeptide sequence selected from the group of sequences set forth in the Sequence Listing, and also include portions of such

polypeptides, wherein such portion of the polypeptide preferably includes at least 30 amino acids and more preferably includes at least 50 amino acids.

“Identity”, as is well understood in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, “identity” also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as determined by the match between strings of such sequences. “Identity” can be readily calculated by known methods including, but not limited to, those described in *Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York (1988); *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*, Griffin, A.M. and Griffin, H.G., eds., Humana Press, New Jersey (1994); *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press (1987); *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., Stockton Press, New York (1991); and Carillo, H., and Lipman, D., *SIAM J Applied Math*, 48:1073 (1988). Methods to determine identity are designed to give the largest match between the sequences tested. Moreover, methods to determine identity are codified in publicly available programs. Computer programs which can be used to determine identity between two sequences include, but are not limited to, GCG (Devereux, J., et al., *Nucleic Acids Research* 12(1):387 (1984); suite of five BLAST programs, three designed for nucleotide sequences queries (BLASTN, BLASTX, and TBLASTX) and two designed for protein sequence queries (BLASTP and TBLASTN) (Coulson, *Trends in Biotechnology*, 12: 76-80 (1994); Birren, et al., *Genome Analysis*, 1: 543-559 (1997)). The BLAST X program is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., et al., NCBI NLM NIH, Bethesda, MD 20894; Altschul, S., et al., *J. Mol. Biol.*, 215:403-410 (1990)). The well known Smith Waterman algorithm can also be used to determine identity.

Parameters for polypeptide sequence comparison typically include the following:

Algorithm: Needleman and Wunsch, *J. Mol. Biol.* 48:443-453 (1970)

Comparison matrix: BLOSSUM62 from Hentikoff and Hentikoff, *Proc. Natl. Acad. Sci USA* 89:10915-10919 (1992)

Gap Penalty: 12

Gap Length Penalty: 4

A program which can be used with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison Wisconsin. The above parameters along with no penalty for end gap are the default parameters for peptide comparisons.

Parameters for polynucleotide sequence comparison include the following:

Algorithm: Needleman and Wunsch, J. Mol. Biol. 48:443-453 (1970)

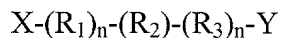
Comparison matrix: matches = +10; mismatches = 0

Gap Penalty: 50

Gap Length Penalty: 3

A program which can be used with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison Wisconsin. The above parameters are the default parameters for nucleic acid comparisons.

The invention also includes polypeptides of the formula:



wherein, at the amino terminus, X is hydrogen, and at the carboxyl terminus, Y is hydrogen or a metal, R₁ and R₃ are any amino acid residue, n is an integer between 1 and 1000, and R₂ is an amino acid sequence of the invention, particularly an amino acid sequence selected from the group set forth in the Sequence Listing and preferably those encoded by the sequences provided in SEQ ID NOs: 2, 4, 6, 9, 12, 17, 19-22, 24-28, 30, 32-35, 37, and 39. In the formula, R₂ is oriented so that its amino terminal residue is at the left, bound to R₁, and its carboxy terminal residue is at the right, bound to R₃. Any stretch of amino acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer.

Polypeptides of the present invention include isolated polypeptides encoded by a polynucleotide comprising a sequence selected from the group of a sequence contained in the Sequence Listing set forth herein .

The polypeptides of the present invention can be mature protein or can be part of a fusion protein.

Fragments and variants of the polypeptides are also considered to be a part of the invention. A fragment is a variant polypeptide which has an amino acid sequence that is entirely the same as part but not all of the amino acid sequence of the previously described polypeptides.

The fragments can be “free-standing” or comprised within a larger polypeptide of which the fragment forms a part or a region, most preferably as a single continuous region. Preferred fragments are biologically active fragments which are those fragments that mediate activities of the polypeptides of the invention, including those with similar activity or improved activity or with a decreased activity. Also included are those fragments that antigenic or immunogenic in an animal, particularly a human.

Variants of the polypeptide also include polypeptides that vary from the sequences set forth in the Sequence Listing by conservative amino acid substitutions, substitution of a residue by another with like characteristics. In general, such substitutions are among Ala, Val, Leu and Ile; between Ser and Thr; between Asp and Glu; between Asn and Gln; between Lys and Arg; or between Phe and Tyr. Particularly preferred are variants in which 5 to 10; 1 to 5; 1 to 3 or one amino acid(s) are substituted, deleted, or added, in any combination.

Variants that are fragments of the polypeptides of the invention can be used to produce the corresponding full length polypeptide by peptide synthesis. Therefore, these variants can be used as intermediates for producing the full-length polypeptides of the invention.

The polynucleotides and polypeptides of the invention can be used, for example, in the transformation of host cells, such as plant host cells, as further discussed herein.

The invention also provides polynucleotides that encode a polypeptide that is a mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids within the mature polypeptide (for example, when the mature form of the protein has more than one polypeptide chain). Such sequences can, for example, play a role in the processing of a protein from a precursor to a mature form, allow protein transport, shorten or lengthen protein half-life, or facilitate manipulation of the protein in assays or production. It is contemplated that cellular enzymes can be used to remove any additional amino acids from the mature protein.

A precursor protein, having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. The inactive precursors generally are activated when the prosequences are removed. Some or all of the prosequences may be removed prior to activation. Such precursor protein are generally called proproteins.

Plant Constructs and Methods of Use

Of particular interest is the use of the nucleotide sequences in recombinant DNA constructs to direct the transcription or transcription and translation (expression) of the prenyltransferase or tocopherol cyclase sequences of the present invention in a host plant cell.

5 The expression constructs generally comprise a promoter functional in a host plant cell operably linked to a nucleic acid sequence encoding a prenyltransferase or tocopherol cyclase of the present invention and a transcriptional termination region functional in a host plant cell.

10 A first nucleic acid sequence is "operably linked" or "operably associated" with a second nucleic acid sequence when the sequences are so arranged that the first nucleic acid sequence affects the function of the second nucleic-acid sequence. Preferably, the two sequences are part of a single contiguous nucleic acid molecule and more preferably are adjacent. For example, a promoter is operably linked to a gene if the promoter regulates or mediates transcription of the gene in a cell.

15 Those skilled in the art will recognize that there are a number of promoters which are functional in plant cells, and have been described in the literature. Chloroplast and plastid specific promoters, chloroplast or plastid functional promoters, and chloroplast or plastid operable promoters are also envisioned.

20 One set of plant functional promoters are constitutive promoters such as the CaMV35S or FMV35S promoters that yield high levels of expression in most plant organs. Enhanced or duplicated versions of the CaMV35S and FMV35S promoters are useful in the practice of this invention (Odell, *et al.* (1985) *Nature* 313:810-812; Rogers, U.S. Patent Number 5,378, 619). In addition, it may also be preferred to bring about expression of the prenyltransferase or tocopherol cyclase gene in specific tissues of the plant, such as leaf, stem, root, tuber, seed, fruit, etc., and the promoter chosen should have the desired tissue and developmental specificity.

25 Of particular interest is the expression of the nucleic acid sequences of the present invention from transcription initiation regions which are preferentially expressed in a plant seed tissue. Examples of such seed preferential transcription initiation sequences include those sequences derived from sequences encoding plant storage protein genes or from genes involved in fatty acid biosynthesis in oilseeds. Examples of such promoters include the 5' regulatory regions from such genes as napin (Kridl *et al.*, *Seed Sci. Res.* 1:209:219 (1991)), phaseolin, zein,
30

soybean trypsin inhibitor, ACP, stearyl-ACP desaturase, soybean α' subunit of β -conglycinin (soy 7s, (Chen *et al.*, *Proc. Natl. Acad. Sci.*, 83:8560-8564 (1986))) and oleosin.

It may be advantageous to direct the localization of proteins conferring prenyltransferase or tocopherol cyclase to a particular subcellular compartment, for example, to the mitochondrion, endoplasmic reticulum, vacuoles, chloroplast or other plastidic compartment. For example, where the genes of interest of the present invention will be targeted to plastids, such as chloroplasts, for expression, the constructs will also employ the use of sequences to direct the gene to the plastid. Such sequences are referred to herein as chloroplast transit peptides (CTP) or plastid transit peptides (PTP). In this manner, where the gene of interest is not directly inserted into the plastid, the expression construct will additionally contain a gene encoding a transit peptide to direct the gene of interest to the plastid. The chloroplast transit peptides may be derived from the gene of interest, or may be derived from a heterologous sequence having a CTP. Such transit peptides are known in the art. See, for example, Von Heijne *et al.* (1991) *Plant Mol. Biol. Rep.* 9:104-126; Clark *et al.* (1989) *J. Biol. Chem.* 264:17544-17550; della-Cioppa *et al.* (1987) *Plant Physiol.* 84:965-968; Romer *et al.* (1993) *Biochem. Biophys. Res Commun.* 196:1414-1421; and, Shah *et al.* (1986) *Science* 233:478-481.

Depending upon the intended use, the constructs may contain the nucleic acid sequence which encodes the entire prenyltransferase or tocopherol cyclase protein, or a portion thereof. For example, where antisense inhibition of a given prenyltransferase or tocopherol cyclase protein is desired, the entire prenyltransferase or tocopherol cyclase sequence is not required. Furthermore, where prenyltransferase or tocopherol cyclase sequences used in constructs are intended for use as probes, it may be advantageous to prepare constructs containing only a particular portion of a prenyltransferase or tocopherol cyclase encoding sequence, for example a sequence which is discovered to encode a highly conserved prenyltransferase or tocopherol cyclase region.

The skilled artisan will recognize that there are various methods for the inhibition of expression of endogenous sequences in a host cell. Such methods include, but are not limited to, antisense suppression (Smith, *et al.* (1988) *Nature* 334:724-726), co-suppression (Napoli, *et al.* (1989) *Plant Cell* 2:279-289), ribozymes (PCT Publication WO 97/10328), and combinations of sense and antisense Waterhouse, *et al.* (1998) *Proc. Natl. Acad. Sci. USA* 95:13959-13964.

Methods for the suppression of endogenous sequences in a host cell typically employ the transcription or transcription and translation of at least a portion of the sequence to be suppressed. Such sequences may be homologous to coding as well as non-coding regions of the endogenous sequence.

5 Regulatory transcript termination regions may be provided in plant expression constructs of this invention as well. Transcript termination regions may be provided by the DNA sequence encoding the prenyltransferase or tocopherol cyclase or a convenient transcription termination region derived from a different gene source, for example, the transcript termination region which is naturally associated with the transcript initiation region. The skilled artisan will recognize that
10 any convenient transcript termination region which is capable of terminating transcription in a plant cell may be employed in the constructs of the present invention.

Alternatively, constructs may be prepared to direct the expression of the prenyltransferase or tocopherol cyclase sequences directly from the host plant cell plastid. Such constructs and methods are known in the art and are generally described, for example, in Svab, *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:8526-8530 and Svab and Maliga (1993) *Proc. Natl. Acad. Sci. USA* 90:913-917 and in U.S. Patent Number 5,693,507.

The prenyltransferase or tocopherol cyclase constructs of the present invention can be used in transformation methods with additional constructs providing for the expression of other nucleic acid sequences encoding proteins involved in the production of tocopherols, or tocopherol precursors such as homogentisic acid and/or phytylpyrophosphate. Nucleic acid sequences encoding proteins involved in the production of homogentisic acid are known in the art, and include but not are limited to, 4-hydroxyphenylpyruvate dioxygenase (HPPD, EC 1.13.11.27) described for example, by Garcia, *et al.* ((1999) *Plant Physiol.* 119(4):1507-1516), mono or bifunctional tyrA (described for example by Xia, *et al.* (1992) *J. Gen Microbiol.* 138:1309-1316, and Hudson, *et al.* (1984) *J. Mol. Biol.* 180:1023-1051), Oxygenase, 4-hydroxyphenylpyruvate di- (9CI), 4-Hydroxyphenylpyruvate dioxygenase; p-Hydroxyphenylpyruvate dioxygenase; p-Hydroxyphenylpyruvate hydroxylase; p-Hydroxyphenylpyruvate oxidase; p-Hydroxyphenylpyruvic acid hydroxylase; p-Hydroxyphenylpyruvic hydroxylase; p-Hydroxyphenylpyruvic oxidase), 4-hydroxyphenylacetate, NAD(P)H:oxygen oxidoreductase (1-hydroxylating); 4-hydroxyphenylacetate 1-monooxygenase,
20
25
30

and the like. In addition, constructs for the expression of nucleic acid sequences encoding proteins involved in the production of phytylpyrophosphate can also be employed with the prenyltransferase or tocopherol cyclase constructs of the present invention. Nucleic acid sequences encoding proteins involved in the production of phytylpyrophosphate are known in the art, and include, but are not limited to geranylgeranylpyrophosphate synthase (GGPPS), geranylgeranylpyrophosphate reductase (GGH), 1-deoxyxylulose-5-phosphate synthase, 1-deoxy-D-xylulose-5-phosphate reductoisomerase, 4-diphosphocytidyl-2-C-methylerythritol synthase, isopentyl pyrophosphate isomerase.

The prenyltransferase or tocopherol cyclase sequences of the present invention find use in the preparation of transformation constructs having a second expression cassette for the expression of additional sequences involved in tocopherol biosynthesis. Additional tocopherol biosynthesis sequences of interest in the present invention include, but are not limited to gamma-tocopherol methyltransferase (Shintani, *et al.* (1998) *Science* 282(5396):2098-2100), tocopherol cyclase, and tocopherol methyltransferase.

A plant cell, tissue, organ, or plant into which the recombinant DNA constructs containing the expression constructs have been introduced is considered transformed, transfected, or transgenic. A transgenic or transformed cell or plant also includes progeny of the cell or plant and progeny produced from a breeding program employing such a transgenic plant as a parent in a cross and exhibiting an altered phenotype resulting from the presence of a prenyltransferase or tocopherol cyclase nucleic acid sequence.

Plant expression or transcription constructs having a prenyltransferase or tocopherol cyclase as the DNA sequence of interest for increased or decreased expression thereof may be employed with a wide variety of plant life, particularly, plant life involved in the production of vegetable oils for edible and industrial uses. Particularly preferred plants for use in the methods of the present invention include, but are not limited to: *Acacia*, alfalfa, aneth, apple, apricot, artichoke, arugula, asparagus, avocado, banana, barley, beans, beet, blackberry, blueberry, broccoli, brussels sprouts, cabbage, canola, cantaloupe, carrot, cassava, cauliflower, celery, cherry, chicory, cilantro, citrus, clementines, coffee, corn, cotton, cucumber, Douglas fir, eggplant, endive, escarole, eucalyptus, fennel, figs, garlic, gourd, grape, grapefruit, honey dew, jicama, kiwifruit, lettuce, leeks, lemon, lime, Loblolly pine, mango, melon, mushroom, nectarine,

nut, oat, oil palm, oil seed rape, okra, onion, orange, an ornamental plant, papaya, parsley, pea, peach, peanut, pear, pepper, persimmon, pine, pineapple, plantain, plum, pomegranate, poplar, potato, pumpkin, quince, radiata pine, radicchio, radish, raspberry, rice, rye, sorghum, Southern pine, soybean, spinach, squash, strawberry, sugarbeet, sugarcane, sunflower, sweet potato,
5 sweetgum, tangerine, tea, tobacco, tomato, triticale, turf, turnip, a vine, watermelon, wheat, yams, and zucchini.

Most especially preferred are temperate oilseed crops. Temperate oilseed crops of interest include, but are not limited to, rapeseed (Canola and High Erucic Acid varieties), sunflower, safflower, cotton, soybean, peanut, coconut and oil palms, and corn. Depending on
10 the method for introducing the recombinant constructs into the host cell, other DNA sequences may be required. Importantly, this invention is applicable to dicotyledons and monocotyledons species alike and will be readily applicable to new and/or improved transformation and regulation techniques.

Of particular interest, is the use of prenyltransferase or tocopherol cyclase constructs in
5 plants to produce plants or plant parts, including, but not limited to leaves, stems, roots, reproductive, and seed, with a modified content of tocopherols in plant parts having transformed plant cells.

For immunological screening, antibodies to the protein can be prepared by injecting rabbits or mice with the purified protein or portion thereof, such methods of preparing antibodies
20 being well known to those in the art. Either monoclonal or polyclonal antibodies can be produced, although typically polyclonal antibodies are more useful for gene isolation. Western analysis may be conducted to determine that a related protein is present in a crude extract of the desired plant species, as determined by cross-reaction with the antibodies to the encoded proteins. When cross-reactivity is observed, genes encoding the related proteins are isolated by
25 screening expression libraries representing the desired plant species. Expression libraries can be constructed in a variety of commercially available vectors, including lambda gt11, as described in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual*, Second Edition (1989) Cold Spring Harbor Laboratory, Cold Spring Harbor, New York).

To confirm the activity and specificity of the proteins encoded by the identified nucleic
30 acid sequences as prenyltransferase or tocopherol cyclase enzymes, *in vitro* assays are performed

in insect cell cultures using baculovirus expression systems. Such baculovirus expression systems are known in the art and are described by Lee, *et al.* U.S. Patent Number 5,348,886, the entirety of which is herein incorporated by reference.

In addition, other expression constructs may be prepared to assay for protein activity utilizing different expression systems. Such expression constructs are transformed into yeast or prokaryotic host and assayed for prenyltransferase or tocopherol cyclase activity. Such expression systems are known in the art and are readily available through commercial sources.

In addition to the sequences described in the present invention, DNA coding sequences useful in the present invention can be derived from algae, fungi, bacteria, mammalian sources, plants, etc. Homology searches in existing databases using signature sequences corresponding to conserved nucleotide and amino acid sequences of prenyltransferase or tocopherol cyclase can be employed to isolate equivalent, related genes from other sources such as plants and microorganisms. Searches in EST databases can also be employed. Furthermore, the use of DNA sequences encoding enzymes functionally enzymatically equivalent to those disclosed herein, wherein such DNA sequences are degenerate equivalents of the nucleic acid sequences disclosed herein in accordance with the degeneracy of the genetic code, is also encompassed by the present invention. Demonstration of the functionality of coding sequences identified by any of these methods can be carried out by complementation of mutants of appropriate organisms, such as *Synechocystis*, *Shewanella*, yeast, *Pseudomonas*, *Rhodobacteria*, etc., that lack specific biochemical reactions, or that have been mutated. The sequences of the DNA coding regions can be optimized by gene resynthesis, based on codon usage, for maximum expression in particular hosts.

For the alteration of tocopherol production in a host cell, a second expression construct can be used in accordance with the present invention. For example, the prenyltransferase or tocopherol cyclase expression construct can be introduced into a host cell in conjunction with a second expression construct having a nucleotide sequence for a protein involved in tocopherol biosynthesis.

The method of transformation in obtaining such transgenic plants is not critical to the instant invention, and various methods of plant transformation are currently available.

Furthermore, as newer methods become available to transform crops, they may also be directly

applied hereunder. For example, many plant species naturally susceptible to *Agrobacterium* infection may be successfully transformed via tripartite or binary vector methods of *Agrobacterium* mediated transformation. In many instances, it will be desirable to have the construct bordered on one or both sides by T-DNA, particularly having the left and right borders, more particularly the right border. This is particularly useful when the construct uses *A. tumefaciens* or *A. rhizogenes* as a mode for transformation, although the T-DNA borders may find use with other modes of transformation. In addition, techniques of microinjection, DNA particle bombardment, and electroporation have been developed which allow for the transformation of various monocot and dicot plant species.

Normally, included with the DNA construct will be a structural gene having the necessary regulatory regions for expression in a host and providing for selection of transformant cells. The gene may provide for resistance to a cytotoxic agent, e.g. antibiotic, heavy metal, toxin, etc., complementation providing prototrophy to an auxotrophic host, viral immunity or the like. Depending upon the number of different host species the expression construct or components thereof are introduced, one or more markers may be employed, where different conditions for selection are used for the different hosts.

Where *Agrobacterium* is used for plant cell transformation, a vector may be used which may be introduced into the *Agrobacterium* host for homologous recombination with T-DNA or the Ti- or Ri-plasmid present in the *Agrobacterium* host. The Ti- or Ri-plasmid containing the T-DNA for recombination may be armed (capable of causing gall formation) or disarmed (incapable of causing gall formation), the latter being permissible, so long as the *vir* genes are present in the transformed *Agrobacterium* host. The armed plasmid can give a mixture of normal plant cells and gall.

In some instances where *Agrobacterium* is used as the vehicle for transforming host plant cells, the expression or transcription construct bordered by the T-DNA border region(s) will be inserted into a broad host range vector capable of replication in *E. coli* and *Agrobacterium*, there being broad host range vectors described in the literature. Commonly used is pRK2 or derivatives thereof. See, for example, Ditta, *et al.*, (*Proc. Nat. Acad. Sci., U.S.A.* (1980) 77:7347-7351) and EPA 0 120 515, which are incorporated herein by reference. Alternatively, one may insert the sequences to be expressed in plant cells into a vector containing separate

replication sequences, one of which stabilizes the vector in *E. coli*, and the other in *Agrobacterium*. See, for example, McBride, *et al.* (*Plant Mol. Biol.* (1990) 14:269-276), wherein the pRiHRI (Jouanin, *et al.*, *Mol. Gen. Genet.* (1985) 201:370-374) origin of replication is utilized and provides for added stability of the plant expression vectors in host *Agrobacterium* cells.

Included with the expression construct and the T-DNA will be one or more markers, which allow for selection of transformed *Agrobacterium* and transformed plant cells. A number of markers have been developed for use with plant cells, such as resistance to chloramphenicol, kanamycin, the aminoglycoside G418, hygromycin, or the like. The particular marker employed is not essential to this invention, one or another marker being preferred depending on the particular host and the manner of construction.

For transformation of plant cells using *Agrobacterium*, explants may be combined and incubated with the transformed *Agrobacterium* for sufficient time for transformation, the bacteria killed, and the plant cells cultured in an appropriate selective medium. Once callus forms, shoot formation can be encouraged by employing the appropriate plant hormones in accordance with known methods and the shoots transferred to rooting medium for regeneration of plants. The plants may then be grown to seed and the seed used to establish repetitive generations and for isolation of vegetable oils.

There are several possible ways to obtain the plant cells of this invention which contain multiple expression constructs. Any means for producing a plant comprising a construct having a DNA sequence encoding the expression construct of the present invention, and at least one other construct having another DNA sequence encoding an enzyme are encompassed by the present invention. For example, the expression construct can be used to transform a plant at the same time as the second construct either by inclusion of both expression constructs in a single transformation vector or by using separate vectors, each of which express desired genes. The second construct can be introduced into a plant which has already been transformed with the prenyltransferase or tocopherol cyclase expression construct, or alternatively, transformed plants, one expressing the prenyltransferase or tocopherol cyclase construct and one expressing the second construct, can be crossed to bring the constructs together in the same plant.

Transgenic plants of the present invention may be produced from tissue culture, and subsequent generations grown from seed. Alternatively, transgenic plants may be grown using apomixis. Apomixis is a genetically controlled method of reproduction in plants where the embryo is formed without union of an egg and a sperm. There are three basic types of apomictic reproduction: 1) apospory where the embryo develops from a chromosomally unreduced egg in an embryo sac derived from the nucleus, 2) diplospory where the embryo develops from an unreduced egg in an embryo sac derived from the megaspore mother cell, and 3) adventitious embryony where the embryo develops directly from a somatic cell. In most forms of apomixis, pseudogamy or fertilization of the polar nuclei to produce endosperm is necessary for seed viability. In apospory, a nurse cultivar can be used as a pollen source for endosperm formation in seeds. The nurse cultivar does not affect the genetics of the aposporous apomictic cultivar since the unreduced egg of the cultivar develops parthenogenetically, but makes possible endosperm production. Apomixis is economically important, especially in transgenic plants, because it causes any genotype, no matter how heterozygous, to breed true. Thus, with apomictic reproduction, heterozygous transgenic plants can maintain their genetic fidelity throughout repeated life cycles. Methods for the production of apomictic plants are known in the art. See, U.S. Patent No. 5,811,636, which is herein incorporated by reference in its entirety.

The nucleic acid sequences of the present invention can be used in constructs to provide for the expression of the sequence in a variety of host cells, both prokaryotic and eukaryotic. Host cells of the present invention preferably include monocotyledonous and dicotyledonous plant cells.

In general, the skilled artisan is familiar with the standard resource materials which describe specific conditions and procedures for the construction, manipulation and isolation of macromolecules (e.g., DNA molecules, plasmids, etc.), generation of recombinant organisms and the screening and isolating of clones, (see for example, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press (1989); Maliga *et al.*, *Methods in Plant Molecular Biology*, Cold Spring Harbor Press (1995), the entirety of which is herein incorporated by reference; Birren *et al.*, *Genome Analysis: Analyzing DNA*, 1, Cold Spring Harbor, New York, the entirety of which is herein incorporated by reference).

Methods for the expression of sequences in insect host cells are known in the art. Baculovirus expression vectors are recombinant insect viruses in which the coding sequence for a chosen foreign gene has been inserted behind a baculovirus promoter in place of the viral gene, e.g., polyhedrin (Smith and Summers, U.S. Pat. No., 4,745,051, the entirety of which is incorporated herein by reference). Baculovirus expression vectors are known in the art, and are described for example in Doerfler, *Curr. Top. Microbiol. Immunol.* 131:51-68 (1968); Luckow and Summers, *Bio/Technology* 6:47-55 (1988a); Miller, *Annual Review of Microbiol.* 42:177-199 (1988); Summers, *Curr. Comm. Molecular Biology*, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1988); Summers and Smith, *A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures*, Texas Ag. Exper. Station Bulletin No. 1555 (1988), the entireties of which is herein incorporated by reference)

Methods for the expression of a nucleic acid sequence of interest in a fungal host cell are known in the art. The fungal host cell may, for example, be a yeast cell or a filamentous fungal cell. Methods for the expression of DNA sequences of interest in yeast cells are generally described in "Guide to yeast genetics and molecular biology", Guthrie and Fink, eds. Methods in enzymology, Academic Press, Inc. Vol 194 (1991) and Gene expression technology", Goeddel ed, Methods in Enzymology, Academic Press, Inc., Vol 185 (1991).

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC, Manassas, VA), such as HeLa cells, Chinese hamster ovary (CHO) cells, baby hamster kidney (BHK) cells and a number of other cell lines. Suitable promoters for mammalian cells are also known in the art and include, but are not limited to, viral promoters such as that from Simian Virus 40 (SV40) (Fiers *et al.*, *Nature* 273:113 (1978), the entirety of which is herein incorporated by reference), Rous sarcoma virus (RSV), adenovirus (ADV) and bovine papilloma virus (BPV). Mammalian cells may also require terminator sequences and poly-A addition sequences. Enhancer sequences which increase expression may also be included and sequences which promote amplification of the gene may also be desirable (for example methotrexate resistance genes).

Vectors suitable for replication in mammalian cells are well known in the art, and may include viral replicons, or sequences which insure integration of the appropriate sequences

encoding epitopes into the host genome. Plasmid vectors that greatly facilitate the construction of recombinant viruses have been described (*see, for example, Mackett et al, J Virol. 49:857 (1984); Chakrabarti et al., Mol. Cell. Biol. 5:3403 (1985); Moss, In: Gene Transfer Vectors For Mammalian Cells (Miller and Calos, eds., Cold Spring Harbor Laboratory, N.Y., p. 10, (1987);*
5 all of which are herein incorporated by reference in their entirety).

The invention also includes plants and plant parts, such as seed, oil and meal derived from seed, and feed and food products processed from plants, which are enriched in tocopherols. Of particular interest is seed oil obtained from transgenic plants where the tocopherol level has been increased as compared to seed oil of a non-transgenic plant.

10 The harvested plant material may be subjected to additional processing to further enrich the tocopherol content. The skilled artisan will recognize that there are many such processes or methods for refining, bleaching and degumming oil. United States Patent Number 5,932,261, issued August 3, 1999, discloses on such process, for the production of a natural carotene rich refined and deodorised oil by subjecting the oil to a pressure of less than 0.060 mbar and to a
5 temperature of less than 200.degree. C. Oil distilled by this process has reduced free fatty acids, yielding a refined, deodorised oil where Vitamin E contained in the feed oil is substantially retained in the processed oil. The teachings of this patent are incorporated herein by reference.

20 The invention now being generally described, it will be more readily understood by reference to the following examples which are included for purposes of illustration only and are not intended to limit the present invention.

EXAMPLES

25 **Example 1:** Identification of Prenyltransferase or tocopherol cyclase Sequences

PSI-BLAST (Altschul, *et al. (1997) Nuc Acid Res 25:3389-3402*) profiles were generated for both the straight chain and aromatic classes of prenyltransferases. To generate the straight chain profile, a prenyl- transferase from *Porphyra purpurea* (Genbank accession 1709766) was
30 used as a query against the NCBI non-redundant protein database. The *E. coli* enzyme involved

in the formation of ubiquinone, ubiA (genbank accession 1790473) was used as a starting sequence to generate the aromatic prenyltransferase profile. These profiles were used to search public and proprietary DNA and protein data bases. In *Arabidopsis* six putative prenyltransferases of the straight-chain class were identified, ATPT1, (SEQ ID NO:9), ATPT7 (SEQ ID NO:10), ATPT8 (SEQ ID NO:11), ATPT9 (SEQ ID NO:13), ATPT10 (SEQ ID NO:14), and ATPT11 (SEQ ID NO:15), and six were identified of the aromatic class, ATPT2 (SEQ ID NO:1), ATPT3 (SEQ ID NO:3), ATPT4 (SEQ ID NO:5), ATPT5 (SEQ ID NO:7), ATPT6 (SEQ ID NO:8), and ATPT12 (SEQ ID NO:16). Additional prenyltransferase sequences from other plants related to the aromatic class of prenyltransferases, such as soy (SEQ ID NOs: 19-23, the deduced amino acid sequence of SEQ ID NO:23 is provided in SEQ ID NO:24) and maize (SEQ ID NOs:25-29, and 31) are also identified. The deduced amino acid sequence of ZMPT5 (SEQ ID NO:29) is provided in SEQ ID NO:30.

Searches are performed on a Silicon Graphics Unix computer using additional Bioaccelerator hardware and GenWeb software supplied by Compugen Ltd. This software and hardware enables the use of the Smith-Waterman algorithm in searching DNA and protein databases using profiles as queries. The program used to query protein databases is profilesearch. This is a search where the query is not a single sequence but a profile based on a multiple alignment of amino acid or nucleic acid sequences. The profile is used to query a sequence data set, i.e., a sequence database. The profile contains all the pertinent information for scoring each position in a sequence, in effect replacing the "scoring matrix" used for the standard query searches. The program used to query nucleotide databases with a protein profile is tprofilesearch. Tprofilesearch searches nucleic acid databases using an amino acid profile query. As the search is running, sequences in the database are translated to amino acid sequences in six reading frames. The output file for tprofilesearch is identical to the output file for profilesearch except for an additional column that indicates the frame in which the best alignment occurred.

The Smith-Waterman algorithm, (Smith and Waterman (1981) *supra*), is used to search for similarities between one sequence from the query and a group of sequences contained in the database. E score values as well as other sequence information, such as conserved peptide sequences are used to identify related sequences.

To obtain the entire coding region corresponding to the *Arabidopsis* prenyltransferase sequences, synthetic oligo-nucleotide primers are designed to amplify the 5' and 3' ends of partial cDNA clones containing prenyltransferase sequences. Primers are designed according to the respective *Arabidopsis* prenyltransferase sequences and used in Rapid Amplification of cDNA Ends (RACE) reactions (Frohman *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:8998-9002) using the Marathon cDNA amplification kit (Clontech Laboratories Inc, Palo Alto, CA).

Amino acid sequence alignments between ATPT2 (SEQ ID NO:2), ATPT3 (SEQ ID NO:4), ATPT4 (SEQ ID NO:6), ATPT8 (SEQ ID NO:12), and ATPT12 (SEQ ID NO:17) are performed using ClustalW (Figure 1), and the percent identity and similarities are provided in Table 1 below.

Table 1:

	ATPT2	ATPT3	ATPT4	ATPT8	ATPT12
ATPT2 % Identity	12	13	11	15	
% similar	25	25	22	32	
% Gap	17	20	20	9	
ATPT3 % Identity		12	6	22	
% similar		29	16	38	
% Gap		20	24	14	
ATPT4 % Identity			9	14	
% similar			18	29	
% Gap			26	19	
ATPT8 % Identity				7	
% similar				19	
% Gap				20	
ATPT12 % Identity					
% similar					
% Gap					

Example 2: Preparation of Prenyl Transferase Expression Constructs

A plasmid containing the napin cassette derived from pCGN3223 (described in USPN 5,639,790, the entirety of which is incorporated herein by reference) was modified to make it more useful for cloning large DNA fragments containing multiple restriction sites, and to allow the cloning of multiple napin fusion genes into plant binary transformation vectors. An adapter comprised of the self annealed oligonucleotide of sequence
5 CGCGATTAAATGGCGCGCCCTGCAGGCGGCCCTGCAGGGCGCGCCATTAAAT
CGCGATTAAATGGCGCGCCCTGCAGGCGGCCCTGCAGGGCGCGCCATTAAAT
(SEQ ID NO:40) was ligated into the cloning vector pBC SK+ (Stratagene) after digestion with the restriction endonuclease BssHII to construct vector pCGN7765. Plasmids pCGN3223 and pCGN7765 were digested with NotI and ligated together. The resultant vector, pCGN7770,
10 contains the pCGN7765 backbone with the napin seed specific expression cassette from pCGN3223.

The cloning cassette, pCGN7787, essentially the same regulatory elements as pCGN7770, with the exception of the napin regulatory regions of pCGN7770 have been replaced with the double CAMV 35S promoter and the tml polyadenylation and transcriptional
5 termination region.

A binary vector for plant transformation, pCGN5139, was constructed from pCGN1558 (McBride and Summerfelt, (1990) Plant Molecular Biology, 14:269-276). The polylinker of pCGN1558 was replaced as a HindIII/Asp718 fragment with a polylinker containing unique restriction endonuclease sites, AscI, PacI, XbaI, SmaI, BamHI, and NotI. The Asp718 and
20 HindIII restriction endonuclease sites are retained in pCGN5139.

A series of turbo binary vectors are constructed to allow for the rapid cloning of DNA sequences into binary vectors containing transcriptional initiation regions (promoters) and transcriptional termination regions.

The plasmid pCGN8618 was constructed by ligating oligonucleotides 5'-
25 TCGAGGATCCGCGGCCGCAAGCTTCCTGCAGG-3' (SEQ ID NO:41) and 5'-
TCGACCTGCAGGAAGCTTGCGGCCGCGGATCC-3' (SEQ ID NO:42) into SalI/XhoI-
digested pCGN7770. A fragment containing the napin promoter, polylinker and napin 3' region was excised from pCGN8618 by digestion with Asp718I; the fragment was blunt-ended by filling in the 5' overhangs with Klenow fragment then ligated into pCGN5139 that had been digested
30 with Asp718I and HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment.

A plasmid containing the insert oriented so that the napin promoter was closest to the blunted Asp718I site of pCGN5139 and the napin 3' was closest to the blunted HindIII site was subjected to sequence analysis to confirm both the insert orientation and the integrity of cloning junctions. The resulting plasmid was designated pCGN8622.

5 The plasmid pCGN8619 was constructed by ligating oligonucleotides 5'-TCGACCTGCAGGAAGCTTGCGGCCGCGGATCC -3' (SEQ ID NO:43) and 5'-TCGAGGATCCGCGGCCGCAAGCTTCCTGCAGG-3' (SEQ ID NO:44) into SalI/XhoI-digested pCGN7770. A fragment containing the napin promoter, polylinker and napin 3' region was removed from pCGN8619 by digestion with Asp718I; the fragment was blunt-ended by
10 filling in the 5' overhangs with Klenow fragment then ligated into pCGN5139 that had been digested with Asp718I and HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment. A plasmid containing the insert oriented so that the napin promoter was closest to the blunted Asp718I site of pCGN5139 and the napin 3' was closest to the blunted HindIII site was subjected to sequence analysis to confirm both the insert orientation and the integrity of cloning
15 junctions. The resulting plasmid was designated pCGN8623.

The plasmid pCGN8620 was constructed by ligating oligonucleotides 5'-TCGAGGATCCGCGGCCGCAAGCTTCCTGCAGGAGCT -3' (SEQ ID NO:45) and 5'-CCTGCAGGAAGCTTGCGGCCGCGGATCC-3' (SEQ ID NO:46) into SalI/SacI-digested pCGN7787. A fragment containing the d35S promoter, polylinker and tml 3' region was
20 removed from pCGN8620 by complete digestion with Asp718I and partial digestion with NotI. The fragment was blunt-ended by filling in the 5' overhangs with Klenow fragment then ligated into pCGN5139 that had been digested with Asp718I and HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment. A plasmid containing the insert oriented so that the d35S promoter was closest to the blunted Asp718I site of pCGN5139 and the tml 3' was closest
25 to the blunted HindIII site was subjected to sequence analysis to confirm both the insert orientation and the integrity of cloning junctions. The resulting plasmid was designated pCGN8624.

The plasmid pCGN8621 was constructed by ligating oligonucleotides 5'-TCGACCTGCAGGAAGCTTGCGGCCGCGGATCCAGCT -3' (SEQ ID NO:47) and 5'-GGATCCGCGGCCGCAAGCTTCCTGCAGG-3' (SEQ ID NO:48) into SalI/SacI-digested
30

pCGN7787. A fragment containing the d35S promoter, polylinker and tml 3' region was removed from pCGN8621 by complete digestion with Asp718I and partial digestion with NotI. The fragment was blunt-ended by filling in the 5' overhangs with Klenow fragment then ligated into pCGN5139 that had been digested with Asp718I and HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment. A plasmid containing the insert oriented so that the d35S promoter was closest to the blunted Asp718I site of pCGN5139 and the tml 3' was closest to the blunted HindIII site was subjected to sequence analysis to confirm both the insert orientation and the integrity of cloning junctions. The resulting plasmid was designated pCGN8625.

The plasmid construct pCGN8640 is a modification of pCGN8624 described above. A 938bp PstI fragment isolated from transposon Tn7 which encodes bacterial spectinomycin and streptomycin resistance (Fling et al. (1985), *Nucleic Acids Research* 13(19):7095-7106), a determinant for E. coli and Agrobacterium selection, was blunt ended with Pfu polymerase. The blunt ended fragment was ligated into pCGN8624 that had been digested with SpeI and blunt ended with Pfu polymerase. The region containing the PstI fragment was sequenced to confirm both the insert orientation and the integrity of cloning junctions.

The spectinomycin resistance marker was introduced into pCGN8622 and pCGN8623 as follows. A 7.7 Kbp AvrII-SnaBI fragment from pCGN8640 was ligated to a 10.9 Kbp AvrII-SnaBI fragment from pCGN8623 or pCGN8622, described above. The resulting plasmids were pCGN8641 and pCGN8643, respectively.

The plasmid pCGN8644 was constructed by ligating oligonucleotides 5'-GATCACCTGCAGGAAGCTTGCGGCCGCGGATCCAATGCA-3' (SEQ ID NO:49) and 5'-TTGGATCCGCGGCCGCAAGCTTCCTGCAGGT-3' (SEQ ID NO:50) into BamHI-PstI digested pCGN8640.

Synthetic oligonucleotides were designed for use in Polymerase Chain Reactions (PCR) to amplify the coding sequences of ATPT2, ATPT3, ATPT4, ATPT8, and ATPT12 for the preparation of expression constructs and are provided in Table 2 below.

Table 2:

Name	Restriction Site	Sequence	SEQ ID NO:
ATPT2	5' NotI	GGATCCGCGGCCGCAACAATGGAGTC TCTGCTCTCTAGTTCT	51

ATPT2	3' SseI	GGATCCTGCAGGTCACCTCAAAAAA	52
		GGTAACAGCAAGT	
ATPT3	5' NotI	GGATCCGCGGCCGCAATGGCGTT	53
		TTTTGGGCTCTCCCGTGTTT	
ATPT3	3' SseI	GGATCCTGCAGGTTATTGAAAACCTT	54
		CTTCCAAGTACAACCT	
ATPT4	5' NotI	GGATCCGCGGCCGCAATGTGGCG	55
		AAGATCTGTTGTT	
ATPT4	3' SseI	GGATCCTGCAGGTCATGGAGAGTAG	56
		AAGGAAGGAGCT	
ATPT8	5' NotI	GGATCCGCGGCCGCAATGGTACT	57
		TGCCGAGGTTCCAAAGCTTGCCCTCT	
ATPT8	3' SseI	GGATCCTGCAGGTCACCTGTTTCTG	58
		GTGATGACTCTAT	
ATPT12	5' NotI	GGATCCGCGGCCGCAATGACTTC	59
		GATTCTCAACACT	
ATPT12	3' SseI	GGATCCTGCAGGTCAGTGTTGCGAT	60
		GCTAATGCCGT	

The coding sequences of ATPT2, ATPT3, ATPT4, ATPT8, and ATPT12 were all amplified using the respective PCR primers shown in Table 2 above and cloned into the TopoTA vector (Invitrogen). Constructs containing the respective prenyltransferase sequences were digested with NotI and Sse8387I and cloned into the turbobinary vectors described above.

The sequence encoding ATPT2 prenyltransferase was cloned in the sense orientation into pCGN8640 to produce the plant transformation construct pCGN10800 (Figure 2). The ATPT2 sequence is under control of the 35S promoter.

The ATPT2 sequence was also cloned in the antisense orientation into the construct pCGN8641 to create pCGN10801 (Figure 3). This construct provides for the antisense expression of the ATPT2 sequence from the napin promoter.

The ATPT2 coding sequence was also cloned in the sense orientation into the vector pCGN8643 to create the plant transformation construct pCGN10822

The ATPT2 coding sequence was also cloned in the antisense orientation into the vector pCGN8644 to create the plant transformation construct pCGN10803 (Figure 4).

The ATPT4 coding sequence was cloned into the vector pCGN864 to create the plant transformation construct pCGN10806 (Figure 5). The ATPT2 coding sequence was cloned into the vector TopoTA™ vector from Invitrogen, to create the plant transformation construct pCGN10807 (Figure 6). The ATPT3 coding sequence was cloned into the TopoTA vector to create

the plant transformation construct pCGN10808 (Figure 7). The ATPT3 coding sequence was cloned in the sense orientation into the vector pCGN8640 to create the plant transformation construct pCGN10809 (Figure 8). The ATPT3 coding sequence was cloned in the antisense orientation into the vector pCGN8641 to create the plant transformation construct pCGN10810 (Figure 9). The ATPT3 coding sequence was cloned into the vector pCGN8643 to create the plant transformation construct pCGN10811 (Figure 10). The ATPT3 coding sequence was cloned into the vector pCGN8644 to create the plant transformation construct pCGN10812 (Figure 11). The ATPT4 coding sequence was cloned into the vector pCGN8640 to create the plant transformation construct pCGN10813 (Figure 12). The ATPT4 coding sequence was cloned into the vector pCGN8641 to create the plant transformation construct pCGN10814 (Figure 13). The ATPT4 coding sequence was cloned into the vector pCGN8643 to create the plant transformation construct pCGN10815 (Figure 14). The ATPT4 coding sequence was cloned in the antisense orientation into the vector pCGN8644 to create the plant transformation construct pCGN10816 (Figure 15). The ATPT8 coding sequence was cloned in the sense orientation into the vector pCGN8643 to create the plant transformation construct pCGN10819 (Figure 17). The ATPT12 coding sequence was cloned into the vector pCGN8640 to create the plant transformation construct pCGN10824 (Figure 18). The ATPT12 coding sequence was cloned into the vector pCGN8643 to create the plant transformation construct pCGN10825 (Figure 19). The ATPT8 coding sequence was cloned into the vector pCGN8640 to create the plant transformation construct pCGN10826 (Figure 20).

Example 3: Plant Transformation with Prenyl Transferase Constructs

Transgenic *Brassica* plants are obtained by *Agrobacterium*-mediated transformation as described by Radke *et al.* (*Theor. Appl. Genet.* (1988) 75:685-694; *Plant Cell Reports* (1992) 11:499-505). Transgenic *Arabidopsis thaliana* plants may be obtained by *Agrobacterium*-mediated transformation as described by Valverkens *et al.*, (*Proc. Nat. Acad. Sci.* (1988) 85:5536-5540), or as described by Bent *et al.* ((1994), *Science* 265:1856-1860), or Bechtold *et al.* ((1993), *C.R.Acad.Sci, Life Sciences* 316:1194-1199). Other plant species may be similarly transformed using related techniques.

Alternatively, microprojectile bombardment methods, such as described by Klein *et al.* (*Bio/Technology* 10:286-291) may also be used to obtain nuclear transformed plants.

Example 4: Identification of Additional Prenyltransferases

Additional BLAST searches were performed using the ATPT2 sequence, a sequence in the class of aromatic prenyltransferases. ESTs, and in some case, full-length coding regions, were identified in proprietary DNA libraries.

Soy full-length homologs to ATPT2 were identified by a combination of BLAST (using ATPT2 protein sequence) and 5' RACE. Two homologs resulted (SEQ ID NO:95 and SEQ ID NO:96). Translated amino acid sequences are provided by SEQ ID NO:97 and SEQ ID NO:98.

A rice est ATPT2 homolog is shown in SEQ ID NO:99 (obtained from BLAST using the wheat ATPT2 homolog).

Other homolog sequences were obtained using ATPT2 and PSI-BLAST, including est sequences from wheat (SEQ ID NO:100), leek (SEQ ID NOs:101 and 102), canola (SEQ ID NO:103), corn (SEQ ID NOs:104, 105 and 106), cotton (SEQ ID NO:107) and tomato (SEQ ID NO:108).

A PSI-Blast profile generated using the *E. coli* ubiA (genbank accession 1790473) sequence was used to analyze the *Synechocystis* genome. This analysis identified 5 open reading frames (ORFs) in the *Synechocystis* genome that were potentially prenyltransferases; slr0926 (annotated as ubiA (4-hydroxybenzoate-octaprenyltransferase, SEQ ID NO:32), slr1899 (annotated as ctaB (cytochrome c oxidase folding protein, SEQ ID NO:33), slr0056 (annotated as g4 (chlorophyll synthase 33 kd subunit, SEQ ID NO:34), slr1518 (annotated as menA (menaquinone biosynthesis protein, SEQ ID NO:35), and slr1736 (annotated as a hypothetical protein of unknown function (SEQ ID NO:36).

4A. *Synechocystis* Knock-outs

To determine the functionality of these ORFs and their involvement, if any, in the biosynthesis of tocopherols, knockout constructs were made to disrupt the ORF identified in *Synechocystis*.

Synthetic oligos were designed to amplify regions from the 5' (5'-TAATGTGTACATTGTCGGCCTC (17365')) (SEQ ID NO:61) and 5'-GCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTCCACAATTCCCCGCACCGTC (1736kanpr1)) (SEQ ID NO:62) and 3' (5'-AGGCTAATAAGCACAAATGGGA (17363')) (SEQ ID NO:63) and 5'-GGTATGAGTCAGCAACACCTTCTTCACGAGGCAGACCTCAGCGGAATTGGTTTAGGTTATCCC (1736kanpr2)) (SEQ ID NO:64) ends of the slr1736 ORF. The 1736kanpr1 and 1736kanpr2 oligos contained 20 bp of homology to the slr1736 ORF with an additional 40 bp of sequence homology to the ends of the kanamycin resistance cassette. Separate PCR steps were completed with these oligos and the products were gel purified and combined with the kanamycin resistance gene from puc4K (Pharmacia) that had been digested with *HincII* and gel purified away from the vector backbone. The combined fragments were allowed to assemble without oligos under the following conditions: 94°C for 1 min, 55°C for 1 min, 72°C for 1 min plus 5 seconds per cycle for 40 cycles using pfu polymerase in 100ul reaction volume (Zhao, H and Arnold (1997) *Nucleic Acids Res.* 25(6):1307-1308). One microliter or five microliters of this assembly reaction was then amplified using 5' and 3' oligos nested within the ends of the ORF fragment, so that the resulting product contained 100-200 bp of the 5' end of the *Synechocystis* gene to be knocked out, the kanamycin resistance cassette, and 100-200 bp of the 3' end of the gene to be knocked out. This PCR product was then cloned into the vector pGemT easy (Promega) to create the construct pMON21681 and used for *Synechocystis* transformation.

Primers were also synthesized for the preparation of *Synechocystis* knockout constructs for the other sequences using the same method as described above, with the following primers. The ubiA 5' sequence was amplified using the primers 5'-GGATCCATGGTTGCCCAAACCCCATC (SEQ ID NO:65) and 5'-GCAATGTAACATCAGAGATTTTGAGACACAACG TGGCTTTGGGTAAGCAACAATGACCGGC (SEQ ID NO:66). The 3' region was amplified using the synthetic oligonucleotide primers 5'-GAATTCTCAAAGCCAGCCCAGTAAC (SEQ ID NO:67) and 5'-GGTATGAGTCAGCAACACCTTCTTCACGAGGCAGACCTCAGCGGGTGCGAAAAGGGTTTCCC (SEQ ID NO:68). The amplification products were combined with the kanamycin resistance gene from puc4K (Pharmacia) that had been digested with *HincII* and gel purified away from the vector

backbone. The annealed fragment was amplified using 5' and 3' oligos nested within the ends of the ORF fragment (5'- CCAGTGGTTTAGGCTGTGTGGTC (SEQ ID NO:69) and 5'- CTGAGTTGGATGTATTGGATC (SEQ ID NO:70)), so that the resulting product contained 100-200 bp of the 5' end of the *Synechocystis* gene to be knocked out, the kanamycin resistance cassette, and 100-200 bp of the 3' end of the gene to be knocked out. This PCR product was then cloned into the vector pGemT easy (Promega) to create the construct pMON21682 and used for *Synechocystis* transformation.

Primers were also synthesized for the preparation of *Synechocystis* knockout constructs for the other sequences using the same method as described above, with the following primers.

The sl11899 5' sequence was amplified using the primers 5'- GGATCCATGGTTACTT CGACAAAAATCC (SEQ ID NO:71) and 5'- GCAATGTAACATCAGAG ATTTTGAGACACAACGTGGCTTTGCTAGGCAACCGCTTAGTAC (SEQ ID NO:72). The 3' region was amplified using the synthetic oligonucleotide primers 5'- GAATTCTTAACCCAACAGTAAAGTTCCC (SEQ ID NO:73) and 5'- GGTATGAGTCAGC AACACCTTCTTCACGAGGCAGACCTCAGCGCCGGCATTGTCTTTTACATG (SEQ ID NO:74). The amplification products were combined with the kanamycin resistance gene from puc4K (Pharmacia) that had been digested with *HincII* and gel purified away from the vector backbone. The annealed fragment was amplified using 5' and 3' oligos nested within the ends of the ORF fragment (5'- GGAACCCTTGCAGCCGCTTC (SEQ ID NO:75) and 5'- GTATGCCCAACTGGTGCAGAGG (SEQ ID NO:76)), so that the resulting product contained 100-200 bp of the 5' end of the *Synechocystis* gene to be knocked out, the kanamycin resistance cassette, and 100-200 bp of the 3' end of the gene to be knocked out. This PCR product was then cloned into the vector pGemT easy (Promega) to create the construct pMON21679 and used for *Synechocystis* transformation.

Primers were also synthesized for the preparation of *Synechocystis* knockout constructs for the other sequences using the same method as described above, with the following primers. The slr0056 5' sequence was amplified using the primers 5'- GGATCCATGTCTGACACACAAAATACCG (SEQ ID NO:77) and 5'- GCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTCGCCAATACCAGCCACC AACAG (SEQ ID NO:78). The 3' region was amplified using the synthetic oligonucleotide

primers 5'- GAATTCTCAAAT CCCCGCATGGCCTAG (SEQ ID NO:79) and 5'-
GGTATGAGTCAGCAACACCTTCTTCACGAGGCAGACCTCAGCGGCCTACGGCTTGGA
CGTGTGGG (SEQ ID NO:80). The amplification products were combined with the kanamycin
resistance gene from puc4K (Pharmacia) that had been digested with *HincII* and gel purified
away from the vector backbone. The annealed fragment was amplified using 5' and 3' oligos
nested within the ends of the ORF fragment (5'- CACTTGGATTCCCCTGATCTG (SEQ ID
NO:81) and 5'- GCAATACCCGCTTGGAACG (SEQ ID NO:82)), so that the resulting
product contained 100-200 bp of the 5' end of the *Synechocystis* gene to be knocked out, the
kanamycin resistance cassette, and 100-200 bp of the 3' end of the gene to be knocked out. This
PCR product was then cloned into the vector pGemT easy (Promega) to create the construct
pMON21677 and used for *Synechocystis* transformation.

Primers were also synthesized for the preparation of *Synechocystis* knockout constructs
for the other sequences using the same method as described above, with the following primers.
The slr1518 5' sequence was amplified using the primers 5'- GGATCCATGACCGAAT
CTTCGCCCCTAGC (SEQ ID NO:83) and 5'-GCAATGTAACATCAGAGATTTGA
GACACAACGTGGC TTTCAATCCTAGGTAGCCGAGGCG (SEQ ID NO:84). The 3' region
was amplified using the synthetic oligonucleotide primers 5'- GAATTCTTAGCCCAGGCC
AGCCCAGCC (SEQ ID NO:85) and 5'- GGTATGAGTCAGCAACACCTTCTTCACGA
GGCAGACCTCAGCGGGGAATTGATTTGTTTAATTACC (SEQ ID NO:86). The
amplification products were combined with the kanamycin resistance gene from puc4K
(Pharmacia) that had been digested with *HincII* and gel purified away from the vector backbone.
The annealed fragment was amplified using 5' and 3' oligos nested within the ends of the ORF
fragment (5'- GCGATCGCCATTATCGCTTGG (SEQ ID NO:87) and 5'-
GCAGACTGGCAATTATCAGTAACG (SEQ ID NO:88)), so that the resulting product
contained 100-200 bp of the 5' end of the *Synechocystis* gene to be knocked out, the kanamycin
resistance cassette, and 100-200 bp of the 3' end of the gene to be knocked out. This PCR
product was then cloned into the vector pGemT easy (Promega) to create the construct
pMON21680 and used for *Synechocystis* transformation.

4B. Transformation of *Synechocystis*

Cells of *Synechocystis* 6803 were grown to a density of approximately 2×10^8 cells per ml and harvested by centrifugation. The cell pellet was re-suspended in fresh BG-11 medium (ATCC Medium 616) at a density of 1×10^9 cells per ml and used immediately for transformation. One-hundred microliters of these cells were mixed with 5 ul of mini prep DNA and incubated with light at 30C for 4 hours. This mixture was then plated onto nylon filters resting on BG-11 agar supplemented with TES pH8 and allowed to grow for 12-18 hours. The filters were then transferred to BG-11 agar + TES + 5ug/ml kanamycin and allowed to grow until colonies appeared within 7-10 days (Packer and Glazer, 1988). Colonies were then picked into BG-11 liquid media containing 5 ug/ml kanamycin and allowed to grow for 5 days. These cells were then transferred to Bg-11 media containing 10ug/ml kanamycin and allowed to grow for 5 days and then transferred to Bg-11 + kanamycin at 25ug/ml and allowed to grow for 5 days. Cells were then harvested for PCR analysis to determine the presence of a disrupted ORF and also for HPLC analysis to determine if the disruption had any effect on tocopherol levels.

PCR analysis of the *Synechocystis* isolates for slr1736 and sl1899 showed complete segregation of the mutant genome, meaning no copies of the wild type genome could be detected in these strains. This suggests that function of the native gene is not essential for cell function. HPLC analysis of these same isolates showed that the sl1899 strain had no detectable reduction in tocopherol levels. However, the strain carrying the knockout for slr1736 produced no detectable levels of tocopherol.

The amino acid sequences for the *Synechocystis* knockouts are compared using ClustalW, and are provided in Table 3 below. Provided are the percent identities, percent similarity, and the percent gap. The alignment of the sequences is provided in Figure 21.

Table 3:

	Slr1736	slr0926	sl1899	slr0056	slr1518
slr1736 %identity		14	12	18	11
%similar		29	30	34	26
%gap		8	7	10	5
slr0926 %identity			20	19	14
%similar			39	32	28

%gap	7	9	4
sll1899 %identity		17	13
%similar		29	29
%gap		12	9
slr0056 %identity			15
%similar			31
%gap			8
slr1518 %identity			
%similar			
%gap			

Amino acid sequence comparisons are performed using various *Arabidopsis* prenyltransferase sequences and the *Synechocystis* sequences. The comparisons are presented in Table 4 below. Provided are the percent identities, percent similarity, and the percent gap. The alignment of the sequences is provided in Figure 22.

Table 4:

	ATPT2	slr1736	ATPT3	slr0926	ATPT4	sll1899	ATPT12	slr0056	ATPT8	slr1518
ATPT2		29	9	9	8	8	12	9	7	9
		46	23	21	20	20	28	23	21	20
		27	13	28	23	29	11	24	25	24
slr1736			9	13	8	12	13	15	8	10
			19	28	19	28	26	33	21	26
			34	12	34	15	26	10	12	10
ATPT3				23	11	14	13	10	5	11
				36	26	26	26	21	14	22
				29	21	31	16	30	30	30
					12	20	17	20	11	14
slr0926					24	37	28	33	24	29
					33	12	25	10	11	9

	18	11	8	6	7
ATPT4	33	23	18	16	19
	28	19	32	32	33
		13	17	10	12
slr1899		24	30	23	26
		27	13	10	11
			52	8	11
ATPT1			66	19	26
2			18	25	23
				9	13
slr0056				23	32
				10	8
					7
ATPT8					23
					7
slr1518					

4C. Phytyl Prenyltransferase Enzyme Assays

[³H] Homogentisic acid in 0.1% H₃PO₄ (specific radioactivity 40 Ci/mmol). Phytyl pyrophosphate was synthesized as described by Joo, *et al.* (1973) *Can J. Biochem.* 51:1527. 2-methyl-6-phytylquinol and 2,3-dimethyl-5-phytylquinol were synthesized as described by Soll, *et al.* (1980) *Phytochemistry* 19:215. Homogentisic acid, α, β, δ, and γ-tocopherol, and tocol, were purchased commercially.

The wild-type strain of *Synechocystis* sp. PCC 6803 was grown in BG11 medium with bubbling air at 30°C under 50 μE.m⁻².s⁻¹ fluorescent light, and 70% relative humidity. The growth medium of slr1736 knock-out (potential PPT) strain of this organism was supplemented with 25

$\mu\text{g mL}^{-1}$ kanamycin. Cells were collected from 0.25 to 1 liter culture by centrifugation at 5000 *g* for 10 min and stored at -80°C .

Total membranes were isolated according to Zak's procedures with some modifications (Zak, *et al.* (1999) *Eur J. Biochem* 261:311). Cells were broken on a French press. Before the French press treatment, the cells were incubated for 1 hour with lysozyme (0.5%, w/v) at 30°C in a medium containing 7 mM EDTA, 5 mM NaCl and 10 mM Hepes-NaOH, pH 7.4. The spheroplasts were collected by centrifugation at 5000 *g* for 10 min and resuspended at 0.1 - 0.5 mg chlorophyll $\cdot\text{mL}^{-1}$ in 20 mM potassium phosphate buffer, pH 7.8. Proper amount of protease inhibitor cocktail and DNAase I from Boehringer Mannheim were added to the solution. French press treatments were performed two to three times at 100 MPa. After breakage, the cell suspension was centrifuged for 10 min at 5000*g* to pellet unbroken cells, and this was followed by centrifugation at 100 000 *g* for 1 hour to collect total membranes. The final pellet was resuspended in a buffer containing 50 mM Tris-HCL and 4 mM MgCl_2 .

Chloroplast pellets were isolated from 250 g of spinach leaves obtained from local markets. Devined leaf sections were cut into grinding buffer (2 l /250 g leaves) containing 2 mM EDTA, 1 mM MgCl_2 , 1 mM MnCl_2 , 0.33 M sorbitol, 0.1% ascorbic acid, and 50 mM Hepes at pH 7.5. The leaves were homogenized for 3 sec three times in a 1-L blender, and filtered through 4 layers of miracloth. The supernatant was then centrifuged at 5000*g* for 6 min. The chloroplast pellets were resuspended in small amount of grinding buffer (Douce,*et al* Methods in Chloroplast Molecular Biology, 239 (1982)

Chloroplasts in pellets can be broken in three ways. Chloroplast pellets were first aliquoted in 1 mg of chlorophyll per tube, centrifuged at 6000 rpm for 2 min in microcentrifuge, and grinding buffer was removed. Two hundred microliters of Triton X-100 buffer (0.1% Triton X-100, 50 mM Tris-HCl pH 7.6 and 4 mM MgCl_2) or swelling buffer (10 mM Tris pH 7.6 and 4 mM MgCl_2) was added to each tube and incubated for $\frac{1}{2}$ hour at 4°C . Then the broken chloroplast pellets were used for the assay immediately. In addition, broken chloroplasts can also be obtained by freezing in liquid nitrogen and stored at -80°C for $\frac{1}{2}$ hour, then used for the assay.

In some cases chloroplast pellets were further purified with 40%/ 80% percoll gradient to obtain intact chloroplasts. The intact chloroplasts were broken with swelling buffer, then either used for assay or further purified for envelope membranes with 20.5%/ 31.8% sucrose density

gradient (Sol, *et al* (1980) *supra*). The membrane fractions were centrifuged at 100 000g for 40 min and resuspended in 50 mM Tris-HCl pH 7.6, 4 mM MgCl₂.

Various amounts of [³H]HGA, 40 to 60 μM unlabelled HGA with specific activity in the range of 0.16 to 4 Ci/mmol were mixed with a proper amount of 1M Tris-NaOH pH 10 to adjust pH to 7.6. HGA was reduced for 4 min with a trace amount of solid NaBH₄. In addition to HGA, standard incubation mixture (final vol 1 mL) contained 50 mM Tris-HCl, pH 7.6, 3-5 mM MgCl₂, and 100 μM phytyl pyrophosphate. The reaction was initiated by addition of *Synechocystis* total membranes, spinach chloroplast pellets, spinach broken chloroplasts, or spinach envelope membranes. The enzyme reaction was carried out for 2 hour at 23°C or 30°C in the dark or light. The reaction is stopped by freezing with liquid nitrogen, and stored at -80°C or directly by extraction.

A constant amount of tocol was added to each assay mixture and reaction products were extracted with a 2 mL mixture of chloroform/methanol (1:2, v/v) to give a monophasic solution. NaCl solution (2 mL; 0.9%) was added with vigorous shaking. This extraction procedure was repeated three times. The organic layer containing the prenylquinones was filtered through a 20 μm filter, evaporated under N₂, and then resuspended in 100 μL of ethanol.

The samples were mainly analyzed by Normal-Phase HPLC method (Isocratic 90% Hexane and 10% Methyl-t-butyl ether), and use a Zorbax silica column, 4.6 x 250 mm. The samples were also analyzed by Reversed-Phase HPLC method (Isocratic 0.1% H₃PO₄ in MeOH), and use a Vydac 201HS54 C18 column; 4.6 x 250 mm coupled with an All-tech C18 guard column. The amount of products were calculated based on the substrate specific radioactivity, and adjusted according to the % recovery based on the amount of internal standard.

The amount of chlorophyll was determined as described in Arnon (1949) *Plant Physiol.* 24:1. Amount of protein was determined by the Bradford method using gamma globulin as a standard (Bradford, (1976) *Anal. Biochem.* 72:248)

Results of the assay demonstrate that 2-Methyl-6-Phytylplastoquinone is not produced in the *Synechocystis* slr1736 knockout preparations. The results of the phytyl prenyltransferase enzyme activity assay for the slr1736 knock out are presented in Figure 23.

4D. Complementation of the slr1736 knockout with ATPT2

In order to determine whether ATPT2 could complement the knockout of slr1736 in *Synechocystis* 6803, a plasmid was constructed to express the ATPT2 sequence from the TAC promoter. A vector, plasmid psl1211, was obtained from the lab of Dr. Himadri Pakrasi of Washington University, and is based on the plasmid RSF1010 which is a broad host range plasmid (Ng W.-O., Zentella R., Wang, Y., Taylor J-S. A., Pakrasi, H.B. 2000. *phrA*, the major photoreactivating factor in the cyanobacterium *Synechocystis* sp. strain PCC 6803 codes for a cyclobutane pyrimidine dimer specific DNA photolyase. *Arch. Microbiol.* (in press)). The ATPT2 gene was isolated from the vector pCGN10817 by PCR using the following primers. ATPT2nco.pr 5'-CCATGGATTTCGAGTAAAGTTGTCGC (SEQ ID NO:89); ATPT2ri.pr- 5'-GAATTCACCTTCAAAAAAGGTAACAG (SEQ ID NO:90). These primers will remove approximately 112 BP from the 5' end of the ATPT2 sequence, which is thought to be the chloroplast transit peptide. These primers will also add an NcoI site at the 5' end and an EcoRI site at the 3' end which can be used for sub-cloning into subsequent vectors. The PCR product from using these primers and pCGN10817 was ligated into pGEM T easy and the resulting vector pMON21689 was confirmed by sequencing using the m13forward and m13reverse primers. The NcoI/EcoRI fragment from pMON21689 was then ligated with the EagI/EcoRI and EagI/NcoI fragments from psl1211 resulting in pMON21690. The plasmid pMON21690 was introduced into the slr1736 *Synechocystis* 6803 KO strain via conjugation. Cells of sl906 (a helper strain) and DH10B cells containing pMON21690 were grown to log phase (O.D. 600=0.4) and 1 ml was harvested by centrifugation. The cell pellets were washed twice with a sterile BG-11 solution and resuspended in 200 ul of BG-11. The following was mixed in a sterile eppendorf tube: 50 ul SL906, 50 ul DH10B cells containing pMON21690, and 100 ul of a fresh culture of the slr1736 *Synechocystis* 6803 KO strain (O.D. 730 = 0.2-0.4). The cell mixture was immediately transferred to a nitrocellulose filter resting on BG-11 and incubated for 24 hours at 30C and 2500 LUX(50 ue) of light. The filter was then transferred to BG-11 supplemented with 10ug/ml Gentamycin and incubated as above for ~5 days. When colonies appeared, they were picked and grown up in liquid BG-11 + Gentamycin 10 ug/ml. (Elhai, J. and Wolk, P. 1988. Conjugal transfer of DNA to Cyanobacteria. *Methods in Enzymology* 167, 747-54) The liquid cultures were then assayed for tocopherols by harvesting 1ml of culture by centrifugation, extracting with ethanol/pyrogallol, and HPLC separation. The slr1736 *Synechocystis* 6803 KO

strain, did not contain any detectable tocopherols, while the slr1736 *Synechocystis* 6803 KO strain transformed with pmon21690 contained detectable alpha tocopherol. A *Synechocystis* 6803 strain transformed with psl1211(vector control) produced alpha tocopherol as well.

5 4E: Additional Evidence of Prenyltransferase Activity

To test the hypothesis that slr1736 or ATPT2 are sufficient as single genes to obtain phytyl prenyltransferase activity, both genes were expressed in SF9 cells and in yeast. When either slr1736 or ATPT2 were expressed in insect cells (Table 5) or in yeast, phytyl prenyltransferase activity was detectable in membrane preparations, whereas membrane preparations of the yeast vector control, or membrane preparations of insect cells did not exhibit phytyl prenyltransferase activity.

Table 5: Phytyl prenyltransferase activity

Enzyme source	Enzyme activity [pmol/mg x h]
slr1736 expressed in SF9 cells	20
ATPT2 expressed in SF9 cells	6
SF9 cell control	< 0.05
<i>Synechocystis</i> 6803	0.25
Spinach chloroplasts	0.20

Example 5: Transgenic Plant Analysis

5A. *Arabidopsis*

Arabidopsis plants transformed with constructs for the sense or antisense expression of the ATPT proteins were analyzed by High Pressure Liquid Chromatography (HPLC) for altered levels of total tocopherols, as well as altered levels of specific tocopherols (alpha, beta, gamma, and delta tocopherol).

Extracts of leaves and seeds were prepared for HPLC as follows. For seed extracts, 10 mg of seed was added to 1 g of microbeads (Biospec) in a sterile microfuge tube to which 500 ul 1% pyrogallol (Sigma Chem)/ethanol was added. The mixture was shaken for 3 minutes in a

mini Beadbeater (Biospec) on “fast” speed. The extract was filtered through a 0.2 um filter into an autosampler tube. The filtered extracts were then used in HPLC analysis described below.

Leaf extracts were prepared by mixing 30-50 mg of leaf tissue with 1 g microbeads and freezing in liquid nitrogen until extraction. For extraction, 500 ul 1% pyrogallol in ethanol was added to the leaf/bead mixture and shaken for 1 minute on a Beadbeater (Biospec) on “fast” speed. The resulting mixture was centrifuged for 4 minutes at 14,000 rpm and filtered as described above prior to HPLC analysis.

HPLC was performed on a Zorbax silica HPLC column (4.6 mm X 250 mm) with a fluorescent detection, an excitation at 290 nm, an emission at 336 nm, and bandpass and slits. Solvent A was hexane and solvent B was methyl-t-butyl ether. The injection volume was 20 ul, the flow rate was 1.5 ml/min, the run time was 12 min (40°C) using the gradient (Table 6):

Table 6:

<u>Time</u>	<u>Solvent A</u>	<u>Solvent B</u>
0 min.	90%	10%
10 min.	90%	10%
11 min.	25%	75%
12 min.	90%	10%

Tocopherol standards in 1% pyrogallol/ ethanol were also run for comparison (alpha tocopherol, gamma tocopherol, beta tocopherol, delta tocopherol, and tocopherol (tocol) (all from Matreya).

Standard curves for alpha, beta, delta, and gamma tocopherol were calculated using Chemstation software. The absolute amount of component x is: Absolute amount of x = $\text{Response}_x \times \text{RF}_x \times \text{dilution factor}$ where Response_x is the area of peak x, RF_x is the response factor for component x ($\text{Amount}_x / \text{Response}_x$) and the dilution factor is 500 ul. The ng/mg tissue is found by: total ng component/mg plant tissue.

Results of the HPLC analysis of seed extracts of transgenic *Arabidopsis* lines containing pMON10822 for the expression of ATPT2 from the napin promoter are provided in Figure 24.

HPLC analysis results of segregating T2 *Arabidopsis* seed tissue expressing the ATPT2 sequence from the napin promoter (pCGN10822) demonstrates an increased level of tocopherols in the seed. Total tocopherol levels are increased as much as 50% over the total tocopherol levels of non-transformed (wild-type) *Arabidopsis* plants (Figure 25). Homozygous progeny from the top 3 lines (T3 seed) have up to a two-fold (100%) increase in total tocopherol levels over control *Arabidopsis* seed (Figure 26.)

Furthermore, increases of particular tocopherols are also increased in transgenic *Arabidopsis* plants expressing the ATPT2 nucleic acid sequence from the napin promoter. Levels of delta tocopherol in these lines are increased greater than 3 fold over the delta tocopherol levels obtained from the seeds of wild type *Arabidopsis* lines. Levels of gamma tocopherol in transgenic *Arabidopsis* lines expressing the ATPT2 nucleic acid sequence are increased as much as about 60% over the levels obtained in the seeds of non-transgenic control lines. Furthermore, levels of alpha tocopherol are increased as much as 3 fold over those obtained from non-transgenic control lines.

Results of the HPLC analysis of seed extracts of transgenic *Arabidopsis* lines containing pCGN10803 for the expression of ATPT2 from the enhanced 35S promoter (antisense orientation) are provided in Figure 25. Two lines were identified that have reduced total tocopherols, up to a ten-fold decrease observed in T3 seed compared to control *Arabidopsis* (Figure 27.)

5B. Canola

Brassica napus, variety SP30021, was transformed with pCGN10822 (napin-ATPT2-napin 3', sense orientation) using *Agrobacterium tumefaciens*-mediated transformation. Flowers of the R0 plants were tagged upon pollination and developing seed was collected at 35 and 45 days after pollination (DAP).

Developing seed was assayed for tocopherol levels, as described above for *Arabidopsis*. Line 10822-1 shows a 20% increase of total tocopherols, compared to the wild-type control, at 45 DAP. Figure 28 shows total tocopherol levels measured in developing canola seed.

Example 6: Sequences to Tocopherol Cyclase

6A. Preparation of the slr1737 Knockout

The *Synechocystis* sp. 6803 slr1737 knockout was constructed by the following method.

The GPS™-1 Genome Priming System (New England Biolabs) was used to insert, by a Tn7

Transposase system, a Kanamycin resistance cassette into *slr1737*. A plasmid from a

5 *Synechocystis* genomic library clone containing 652 base pairs of the targeted orf (*Synechocystis* genome base pairs 1324051 – 1324703; the predicted orf base pairs 1323672 – 1324763, as

annotated by Cyanobase) was used as target DNA. The reaction was performed according to the manufacturers protocol. The reaction mixture was then transformed into *E. coli* DH10B

electrocompetant cells and plated. Colonies from this transformation were then screened for

10 transposon insertions into the target sequence by amplifying with M13 Forward and Reverse

Universal primers, yielding a product of 652 base pairs plus ~1700 base pairs, the size of the

transposon kanamycin cassette, for a total fragment size of ~2300 base pairs. After this

determination, it was then necessary to determine the approximate location of the insertion within the targeted orf, as 100 base pairs of orf sequence was estimated as necessary for efficient

5 homologous recombination in *Synechocystis*. This was accomplished through amplification

reactions using either of the primers to the ends of the transposon, Primer S (5' end) or N (3'

end), in combination with either a M13 Forward or Reverse primer. That is, four different primer

combinations were used to map each potential knockout construct: Primer S – M13 Forward,

Primer S – M13 Reverse, Primer N – M13 Forward, Primer N – M13 Reverse. The construct

20 used to transform *Synechocystis* and knockout slr1737 was determined to consist of a

approximately 150 base pairs of slr1737 sequence on the 5' side of the transposon insertion and

approximately 500 base pairs on the 3' side, with the transcription of the orf and kanamycin

cassette in the same direction. The nucleic acid sequence of slr1737 is provided in SEQ ID

NO:38 the deduced amino acid sequence is provided in SEQ ID NO:39.

25 Cells of *Synechocystis* 6803 were grown to a density of $\sim 2 \times 10^8$ cells per ml and

harvested by centrifugation. The cell pellet was re-suspended in fresh BG-11 medium at a

density of 1×10^9 cells per ml and used immediately for transformation. 100 ul of these cells were

mixed with 5 ul of mini prep DNA and incubated with light at 30C for 4 hours. This mixture

was then plated onto nylon filters resting on BG-11 agar supplemented with TES pH8 and

30 allowed to grow for 12-18 hours. The filters were then transferred to BG-11 agar + TES +

5ug/ml kanamycin and allowed to grow until colonies appeared within 7-10 days (Packer and Glazer, 1988). Colonies were then picked into BG-11 liquid media containing 5 ug/ml kanamycin and allowed to grow for 5 days. These cells were then transferred to Bg-11 media containing 10ug/ml kanamycin and allowed to grow for 5 days and then transferred to Bg-11 + kanamycin at 25ug/ml and allowed to grow for 5 days. Cells were then harvested for PCR analysis to determine the presence of a disrupted ORF and also for HPLC analysis to determine if the disruption had any effect on tocopherol levels.

PCR analysis of the *Synechocystis* isolates, using primers to the ends of the *slr1737* orf, showed complete segregation of the mutant genome, meaning no copies of the wild type genome could be detected in these strains. This suggests that function of the native gene is not essential for cell function. HPLC analysis of the strain carrying the knockout for *slr1737* produced no detectable levels of tocopherol.

6B. The relation of *slr1737* and *slr1736*

The *slr1737* gene occurs in *Synechocystis* downstream and in the same orientation as *slr1736*, the phytyl prenyltransferase. In bacteria this proximity often indicates an operon structure and therefore an expression pattern that is linked in all genes belonging to this operon. Occasionally such operons contain several genes that are required to constitute one enzyme. To confirm that *slr1737* is not required for phytyl prenyltransferase activity, phytyl prenyltransferase was measured in extracts from the *Synechocystis* *slr1737* knockout mutant. Figure 29 shows that extracts from the *Synechocystis* *slr1737* knockout mutant still contain phytyl prenyltransferase activity. The molecular organization of genes in *Synechocystis* 6803 is shown in A. Figures B and C show HPLC traces (normal phase HPLC) of reaction products obtained with membrane preparations from *Synechocystis* wild type and *slr1737*⁻ membrane preparations, respectively.

The fact that *slr1737* is not required for the PPT activity provides additional data that ATPT2 and *slr1736* encode phytyl prenyltransferases.

6C *Synechocystis* Knockouts

Synechocystis 6803 wild type and *Synechocystis* *slr1737* knockout mutant were grown photoautotrophically. Cells from a 20 ml culture of the late logarithmic growth phase were

harvested and extracted with ethanol. Extracts were separated by isocratic normal-phase HPLC using a Hexane/Methyl-t-butyl ether (95/5) and a Zorbax silica column, 4.6 x 250 mm. Tocopherols and tocopherol intermediates were detected by fluorescence (excitement 290 nm, emission 336 nm) (Figure 30).

5 Extracts of *Synechocystis* 6803 contained a clear signal of alpha-tocopherol. 2,3-Dimethyl-5-phytylplastoquinol was below the limit of detection in extracts from the *Synechocystis* wild type (C). In contrast, extracts from the *Synechocystis* slr1737 knockout mutant did not contain alpha-tocopherol, but contained 2,3-dimethyl-5-phytylplastoquinol (D), indicating that the interruption of slr1737 has resulted in a block of the 2,3-dimethyl-5-phytylplastoquinol cyclase reaction.

10 Chromatograms of standard compounds alpha, beta, gamma, delta-tocopherol and 2,3-dimethyl-5-phytylplastoquinol are shown in A and B. Chromatograms of extracts from *Synechocystis* wild type and the *Synechocystis* slr1737 knockout mutant are shown in C and D, respectively. Abbreviations: 2,3-DMPQ, 2,3-dimethyl-5-phytylplastoquinol.

6D. Incubation with Lysozyme treated *Synechocystis*

20 *Synechocystis* 6803 wild type and slr1737 knockout mutant cells from the late logarithmic growth phase (approximately 1g wet cells per experiment in a total volume of 3 ml) were treated with Lysozyme and subsequently incubated with S-adenosylmethionine, and phytylpyrophosphate, plus radiolabelled homogentisic acid. After 17h incubation in the dark at room temperature the samples were extracted with 6 ml chloroform / methanol (1/2 v/v). Phase separation was obtained by the addition of 6 ml 0.9% NaCl solution. This procedure was repeated three times. Under these conditions 2,3-dimethyl-5-phytylplastoquinol is oxidized to form 2,3-dimethyl-5-phytylplastoquinone.

25 The extracts were analyzed by normal phase and reverse phase HPLC. Using extracts from wild type *Synechocystis* cells radiolabelled gamma-tocopherol and traces of radiolabelled 2,3-dimethyl-5-phytylplastoquinone were detected. When extracts from the slr1737 knockout mutant were analyzed, only radiolabelled 2,3-dimethyl-5-phytylplastoquinone was detectable. The amount of 2,3-dimethyl-5-phytylplastoquinone was significantly increased compared to wild type extracts. Heat treated samples of the wild type and the slr1737 knockout mutant did not

30

produce radiolabelled 2,3-dimethyl-5-phytylplastoquinone, nor radiolabelled tocopherols. These results further support the role of the slr1737 expression product in the cyclization of 2,3-dimethyl-5-phytylplastoquinol.

5 6E. *Arabidopsis* Homologue to slr1737

An *Arabidopsis* homologue to slr1737 was identified from a BLASTALL search using *Synechocystis* sp 6803 gene slr1737 as the query, in both public and proprietary databases. SEQ ID NO:109 and SEQ ID NO:110 are the DNA and translated amino acid sequences, respectively, of the *Arabidopsis* homologue to slr1737. The start is found at the ATG at base 56 in SEQ ID NO:109.

The sequences obtained for the homologue from the proprietary database differs from the public database (F4D11.30, BAC AL022537), in having a start site 471 base pairs upstream of the start identified in the public sequence. A comparison of the public and proprietary sequences is provided in Figure 31. The correct start correlates within the public database sequence is at 12080, while the public sequence start is given as being at 11609.

Attempts to amplify a slr1737 homologue were unsuccessful using primers designed from the public database, while amplification of the gene was accomplished with primers obtained from SEQ ID NO:109.

Analysis of the protein sequence to identify transit peptide sequence predicted two potential cleavage sites, one between amino acids 48 and 49, and the other between amino acids 98 and 99.

6F. slr1737 Protein Information

The slr1737 orf comprises 363 amino acid residues and has a predicted MW of 41kDa (SEQ ID NO: 39). Hydropathic analysis indicates the protein is hydrophilic (Figure 32).

The *Arabidopsis* homologue to slr1737 (SEQ ID xx) comprises 488 amino acid residues, has a predicted MW of 55kDa, and has a putative transit peptide sequence comprising the first 98 amino acids. The predicted MW of the mature form of the *Arabidopsis* homologue is 44kDa. The hydropathic plot for the *Arabidopsis* homologue also reveals that it is hydrophilic (Figure 33). Further blast analysis of the *Arabidopsis* homologue reveals limited sequence identity (25 %

sequence identity) with the beta-subunit of respiratory nitrate reductase. Based on the sequence identity to nitrate reductase, it suggests the slr1737 orf is an enzyme that likely involves general acid catalysis mechanism.

Investigation of known enzymes involved in tocopherol metabolism indicated that the best candidate corresponding to the general acid mechanism is the tocopherol cyclase. There are many known examples of cyclases including, tocopherol cyclase, chalcone isomerase, lycopene cyclase, and aristolochene synthase. By further examination of the microscopic catalytic mechanism of phytoplastoquinol cyclization, as an example, chalcone isomerase has a catalytic mechanism most similar to tocopherol cyclase. (Figure 34).

Multiple sequence alignment was performed between slr1737, slr1737 *Arabidopsis* homologue and the *Arabidopsis* chalcone isomerase (Genbank:P41088) (Figure 35). 65% of the conserved residues among the three enzymes are strictly conserved within the known chalcone isomerases. The crystal structure of alfalfa chalcone isomerase has been solved (Jez, Joseph M., Bowman, Marianne E., Dixon, Richard A., and Noel, Joseph P. (2000) "Structure and mechanism of the evolutionarily unique plant enzyme chalcone isomerase". *Nature Structural Biology* 7: 786-791.) It has been demonstrated tyrosine (Y) 106 of the alfalfa chalcone isomerase serves as the general acid during cyclization reaction (Genbank: P28012). The equivalent residue in slr1737 and the slr1737 *Arabidopsis* homolog is lysine (K), which is an excellent catalytic residue as general acid.

The information available from partial purification of tocopherol cyclase from *Chlorella protothecoides* (U.S. Patent No. 5,432,069), *i.e.*, described as being glycine rich, water soluble and with a predicted MW of 48-50kDa, is consistent with the protein informatics information obtained for the slr1737 and the *Arabidopsis* slr1737 homologue.

All publications and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claim.

CLAIMS

What is claimed is:

1. An isolated nucleic acid sequence encoding a tocopherol cyclase.
2. An isolated nucleic acid sequence according to Claim 1, wherein said tocopherol cyclase is active in the cyclization of 2,3-dimethyl-5-phytylplastoquinol to tocopherol.
3. An isolated nucleic acid sequence according to Claim 1, wherein said tocopherol cyclase is active in the cyclization of 2,3-dimethyl-5-geranylgeranylplastoquinol to tocotrienol.
4. An isolated DNA sequence according to Claim 1, wherein said nucleic acid sequence is isolated from a eukaryotic cell source.
5. An isolated DNA sequence according to Claim 4, wherein said eukaryotic cell source is selected from the group consisting of mammalian, nematode, fungal, and plant cells.
6. The DNA encoding sequence of Claim 5 wherein said tocopherol cyclase protein is from *Arabidopsis*.
7. The DNA encoding sequence of Claim 6 wherein said tocopherol cyclase protein is encoded by a sequence of SEQ ID NO:109.
8. The DNA encoding sequence of Claim 7 wherein said tocopherol cyclase protein has an amino acid sequence of SEQ ID NO:110.
9. The DNA encoding sequence of Claim 4 wherein said tocopherol cyclase protein is from a source selected from the group consisting of *Arabidopsis*, soybean, corn, rice, wheat, leek canola, , leek, cotton, and tomato.
10. An isolated DNA sequence according to Claim 4, wherein said prokaryotic source is a *Synechocystis* sp.
11. The DNA encoding sequence of Claim 10 wherein said tocopherol cyclase protein is encoded by a sequence of SEQ ID NO:38.
12. The DNA encoding sequence of Claim 10 wherein said tocopherol cyclase protein has an amino acid sequence of SEQ ID NO:39.
13. A nucleic acid construct comprising as operably linked components, a transcriptional initiation region functional in a host cell, a nucleic acid sequence encoding a tocopherol cyclase, and a transcriptional termination region.

14. A nucleic acid construct according to Claim 13, wherein said nucleic acid sequence encoding tocopherol cyclase is obtained from an organism selected from the group consisting of a eukaryotic organism and a prokaryotic organism.

15. A nucleic acid construct according to Claim 14, wherein said nucleic acid sequence encoding tocopherol cyclase is obtained from a plant source.

16. A nucleic acid construct according to Claim 15, wherein said nucleic acid sequence encoding tocopherol cyclase is obtained from a source selected from the group consisting of *Arabidopsis*, soybean, corn, rice, wheat, leek canola, , leek, cotton, and tomato.

17. A nucleic acid construct according to Claim 13, wherein said nucleic acid sequence encoding tocopherol cyclase is obtained from a *Synechocystis* sp.

18. A plant cell comprising the construct of 13.

19. A plant comprising a cell of Claim 18.

20 A feed composition produced from a plant according to Claim 19.

21. A seed comprising a cell of Claim 18.

22 Oil obtained from a seed of Claim 21.

23. A natural tocopherol rich refined and deodorised oil which has been produced by a method of treating an oil according to Claim 22 by distilling under low pressure and high temperature, wherein said refined oil has reduced free fatty acids and a substantial percentage of tocopherol present in the pretreated oil.

24. A refined oil according to claim 23, wherein the pretreated oil is crude or pre-treated soybean oil.

25. A refined oil according to claim 23, wherein the refined oil is degummed and bleached.

26. A method for the alteration of the isoprenoid content in a host cell, said method comprising; transforming said host cell with a construct comprising as operably linked components, a transcriptional initiation region functional in a host cell, a nucleic acid sequence encoding tocopherol cyclase, and a transcriptional termination region,

wherein said isoprenoid compound selected from the group of tocopherols and tocotrienols .

27. The method according to Claim 26, wherein said host cell is selected from the group consisting of a prokaryotic cell and a eukaryotic cell.

28. The method according to Claim 27, wherein said prokaryotic cell is a *Synechocystis* sp.

29. The method according to Claim 27, wherein said eukaryotic cell is a plant cell.

30. The method according to Claim 29, wherein said plant cell is obtained from a plant selected from the group consisting of *Arabidopsis*, soybean, corn, rice, wheat, leek canola, , leek, cotton, and tomato.

31. A method for producing an isoprenoid compound of interest in a host cell, said method comprising obtaining a transformed host cell, said host cell having and expressing in its genome:

a construct having a DNA sequence encoding a tocopherol cyclase operably linked to a transcriptional initiation region functional in a host cell,

wherein said isoprenoid compound selected from the group of tocopherols and tocotrienols.

32. The method according to Claim 31, wherein said host cell is selected from the group consisting of a prokaryotic cell and a eukaryotic cell.

33. The method according to Claim 32, wherein said prokaryotic cell is a *Synechocystis* sp.

34. The method according to Claim 32, wherein said eukaryotic cell is a plant cell.

35. The method according to Claim 34, wherein said plant cell is obtained from a plant selected from the group consisting wherein said compound selected from the group of *Arabidopsis*, soybean, corn, rice, wheat, leek canola, , leek, cotton, and tomato.

36. A method for increasing the biosynthetic flux in a host cell toward production of an isoprenoid compound, said method comprising;

transforming said host cell with a construct comprising as operably linked components, a transcriptional initiation region functional in a host cell, a DNA encoding a tocopherol cyclase, and a transcriptional termination region,

wherein said isoprenoid compound selected from the group of tocopherols and tocotrienols,.

37. The method according to Claim 36, wherein said host cell is selected from the group consisting of a prokaryotic cell and a eukaryotic cell.

38. The method according to Claim 37, wherein said prokaryotic cell is a *Synechocystis* sp.

39. The method according to Claim 37, wherein said eukaryotic cell is a plant cell.

40. The method according to Claim 39, wherein said plant cell is obtained from a plant selected from the group consisting *Arabidopsis*, soybean, corn, rice, wheat, leek canola, , leek, cotton, and tomato.

41. The method according to Claim 39, wherein said transcriptional initiation region is a seed-specific promoter.

5

**NUCLEIC ACID SEQUENCES TO PROTEINS INVOLVED IN TOCOPHEROL
SYNTHESIS**

5

ABSTRACT

Nucleic acid sequences and methods are provided for producing plants and seeds having altered tocopherol content and compositions. The methods find particular use in increasing the tocopherol levels in plants, and in providing desirable tocopherol compositions in a host plant cell.

10

```

ATPT2 : -----ME-----TSSSSIVSAGC-----FCWKKN-----LKHSILKRLRCD-----SKVAKPK-----FR-----NNTLRP-----DGOG : 59
ATPT3 : MAFGLSRVRRLEKSSVSEPPSSSALLLOSHKSLNPTVTHYNTPLKCYSPMNDNYOMSKGRLEHOKFGVGMNYICGMSSSS : 90
ATPT4 : -----MMRR-----VYRFSSSTSVSSSLPNPLIPNSRL-----CAVNSLQPPVSTESAKRGITGV-----SSD-----ANFA-----TAWA : 69
ATPT8 : -----MVLAEVPKIAS-----AAEFER-----GVGCKR-----SLLIMATA-----LN-----VPLPE-----ALIG : 48
ATPT12 : -----MT-----BNVSTHSSSVTSVDRVGLSHRN-----SDSVEERRR-----SGFSLIYESPCR-----RFV-----VAAE-----FDPD : 64
s 6
r

ATPT2 : 100 * 120 * 140 * 160 * 180
ATPT3 : SSLILY-----KHKSRER-----NATAQPEAFDSNSKQSFDS-----DAFYFRS-----BHVTIGVLSTLSVSLAVEKSDISPLFTGLEA : 141
ATPT4 : SVLEGEKKDKDEKSDGVVKKASWIDLYLPEVRGAYAKLAR-----BKPIGTWLLAMFCMWSIAL-----AADPGLPSF-----K-----YMALFG : 171
ATPT8 : AATAVAT-----TG-EISSR-----AALAGBHHYARCYWEISK-AKASMLVATS-----GTGALGT-----GNAATSGPL-----C-----YTA : 137
ATPT12 : ESTDIV-----SELVRQKGLAEITEMHVASLHDVL-----DDAPTRGVGS-----LNVVGNKMSVLAGDLILS-----RAAG : 117
ATPT12 : KVKSQTP-----DKAPAGSSS-----NDLQIKG-----ASQETNKKWIKIRLOTKRY-TWP-----PLVNGVCGAASGNHWTPE-----VAKSTIL : 140
f c

ATPT2 : * 200 * 220 * 240 * 260 *
ATPT3 : VAAALMNIYIVG-----CHSDVYELKVNKP-----YLILAS-----EYVNTGALIASFSIMS-----FWGVIYGSMPLEWA-----LFVSEMLGTAYSINL : 224
ATPT4 : AL-----LKGAGCT-----DILDDOH-----TKVDTRKRLIAS-----LILPFOG-----GLLOLIG-----IGLLOLNYS-----RVLGASSLIVESY : 248
ATPT8 : GT-----MMIASANS-----OIFEISN-----SKMKRTMLRLES-----RQVPHAPMATIAGASGACLASKT-----MLAG-----LASANIVLYAYEYTP : 219
ATPT12 : A-----AALKNTEVVALATAVEHLVTGE-----MEHISTEORSMDYOKTYUKT-----ASTLSCK-----AAVADGQIAVAV : 190
ATPT12 : NSGPCLFGTYOTLQWYBDRDAINEP-----YKIPSALEPEVITQWVILIG-----LGAGLEVWAGHTTPVIFYIALGGLSYKSA : 227
n d p6 Sg

ATPT2 : 280 * 300 * 320 * 340 * 360
ATPT3 : PLLRMKREFALPAMCLIAVRAITVO-----AFLHIOTHVE-----GPILFTRPLIFAFMSFEES-----VIALFKDIPDI-----G-----KI : 299
ATPT4 : P-----LMKRFTEWFOAFLGL-----INMGALG-----L-----VKGIAPSIHP-----LYLSGYCWT-----VYDITYAHODK-----D-----VK : 314
ATPT8 : LKOLHPINTWGAUV-----GAIPL-----LGA-----ASGOSYNSLIPALYFWQJPHF-----AAHLACRNDYAAGGYKMLSLF-----PSGKRRIA : 300
ATPT12 : LAFEYGRNIGAFOLI-----DDILDEIGTS-----SLGGSLSDRHGVIAPILFAMEEPQIREVVDQVK-----DP-----RN : 259
ATPT12 : PPIKIKONGN-----GNFA-----LG-----ASYIS-----FWAGQ-----HFGITPDIYVY-----TLLYSAG-----GGAIVNDEKSY-----G-----RA : 294
a g e D

ATPT2 : * 380 * 400 * 420 * 440 *
ATPT3 : FGLRS-----FSVTLIG-----KRVFTC-----VTELOMAYAVAILVGLJUSPFMSK-----YISVGHVLAATTIMARAKSYDLSSKTEITSY : 375
ATPT4 : GOKS-----TALRFGD-----NKLUTGFGYISIGFAASGESADLGMWYAS-----IAASGOGWIOIGADLSSGADCSKRVSNKMW : 382
ATPT8 : ALRNCFYMBLGLAYDWGLTSS-----FOLESTLTLIAAATAFSYRDRKMHAKRMFHASY-----ELVEYMSGILHRYVSNINDOOLVEEAGL : 390
ATPT12 : DAL-----EYLKSK-----GLQ-----RARELA-----EHANLAAAGISIPET-----DNEDKRSRALIDLTHRVITRNK : 321
ATPT12 : GLDS-----LVAFC-----EYAKTIG-----VGAIDITQ-----VYAGVLLASGKEYALA-----LVALLIPQYVFOEKYFLKDPVKYDVKYQASAOPE : 373
w

ATPT2 : 460 * 480 *
ATPT3 : MEMLKLFYAE-----YLLLPFLK----- : 393
ATPT4 : GAIFFSGVVLG-----RSFQ----- : 407
ATPT8 : TNSGGEVKTQRRKKVAQEPVAYASNAAPFLPAPSTYSP : 431
ATPT12 : -I-----LGIPIYA-----LASOH----- : 387

```

Figure 1

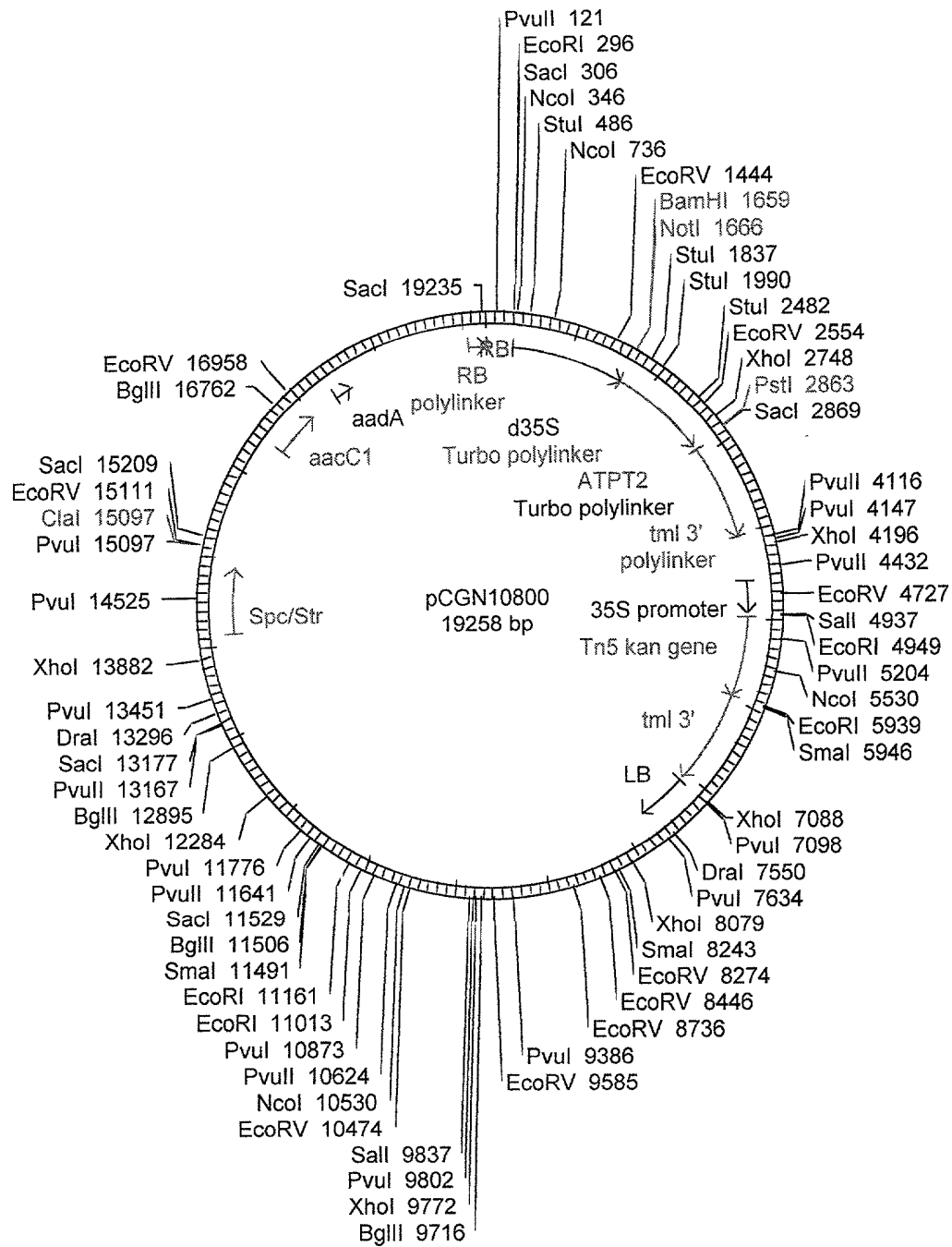


Figure 2

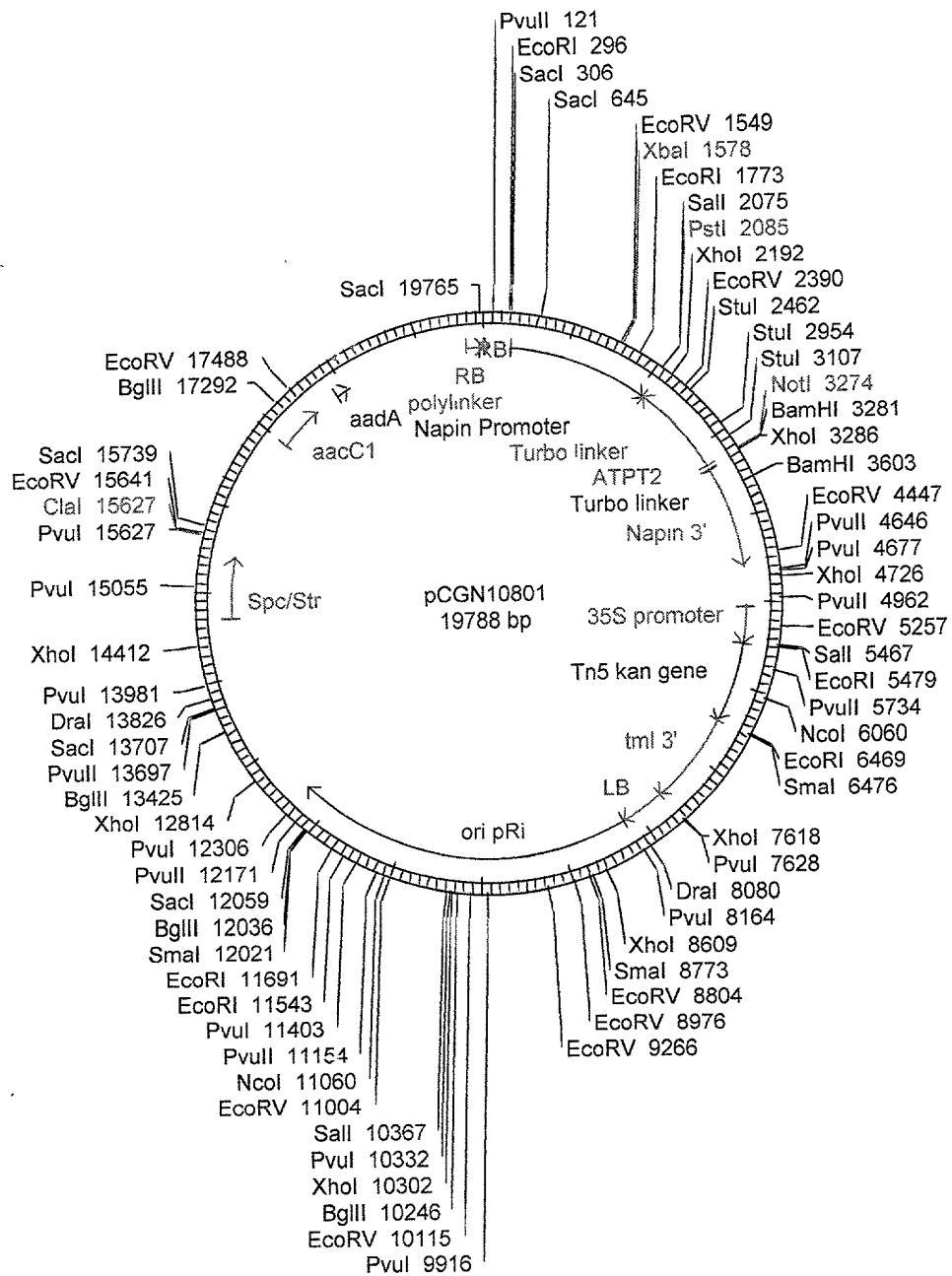


Figure 3

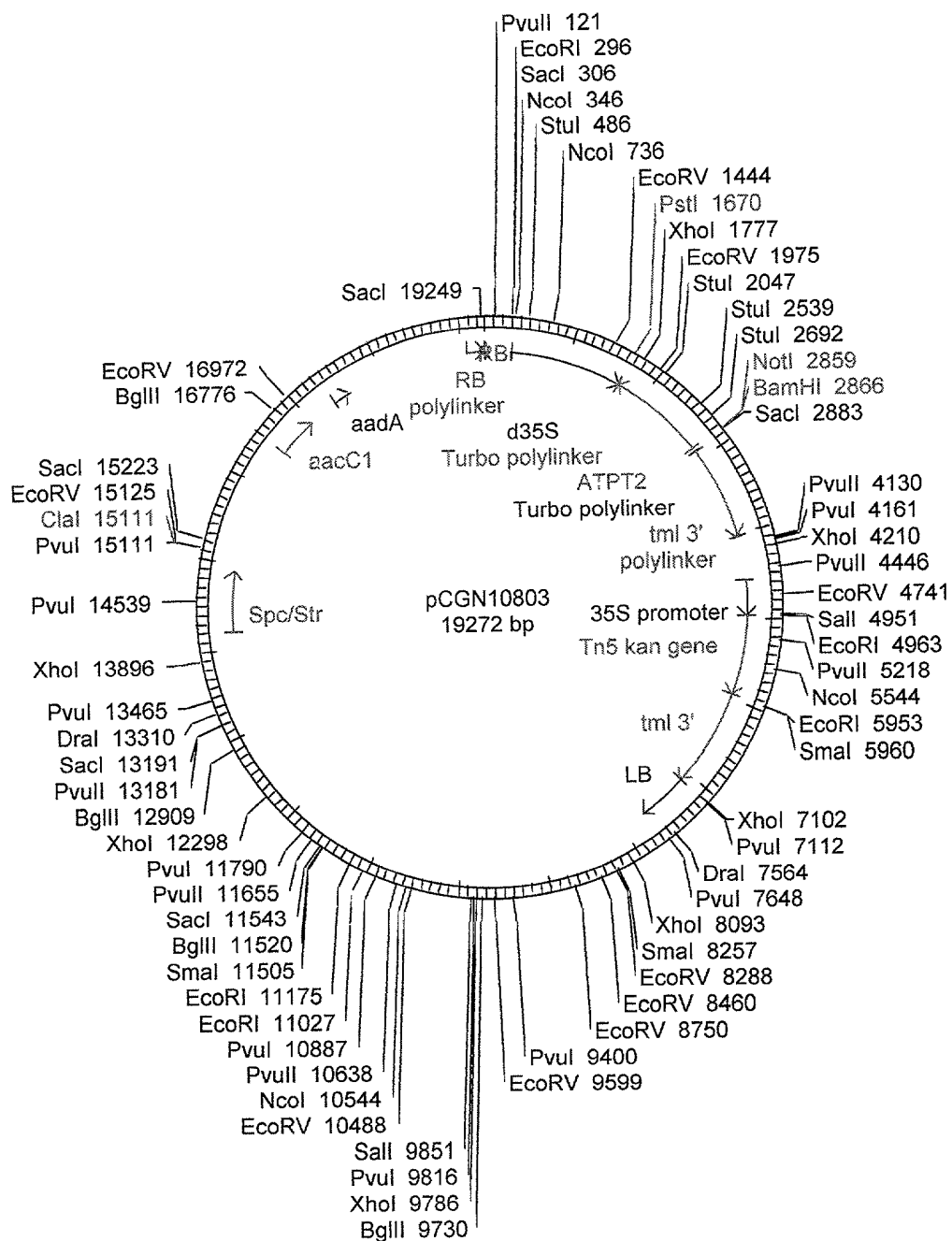


Figure 4

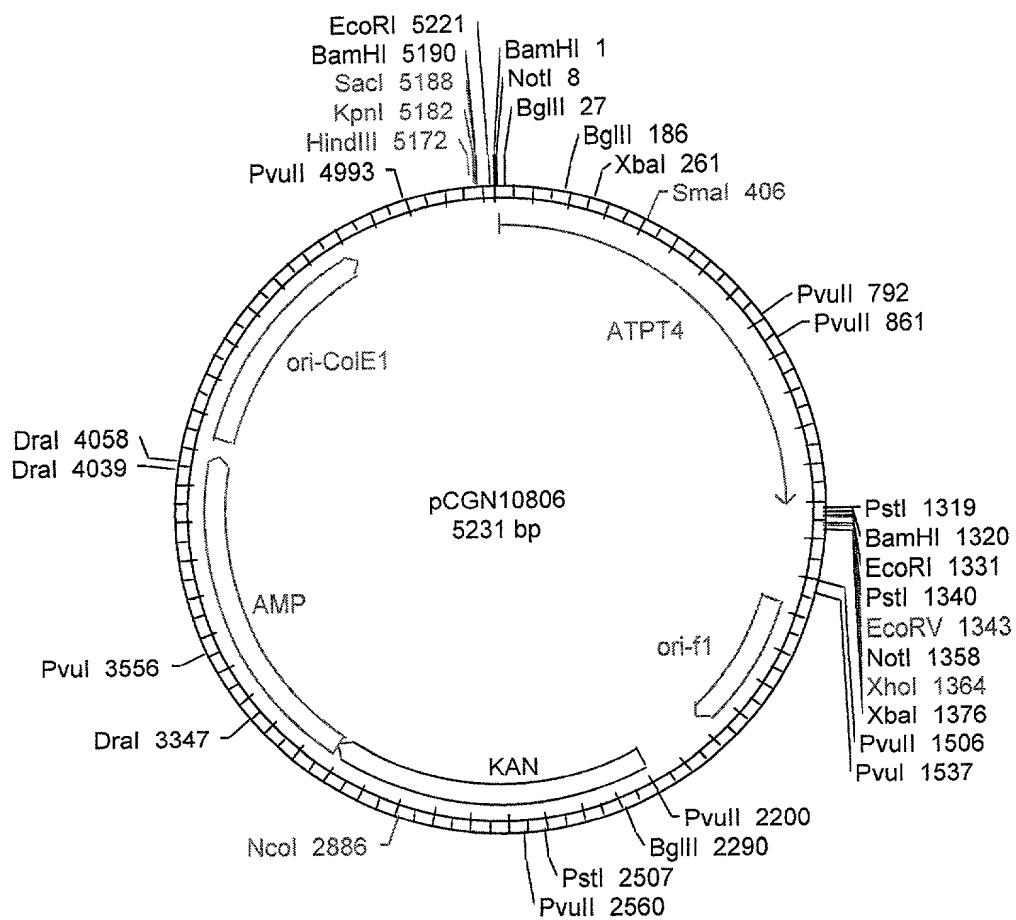


Figure 5

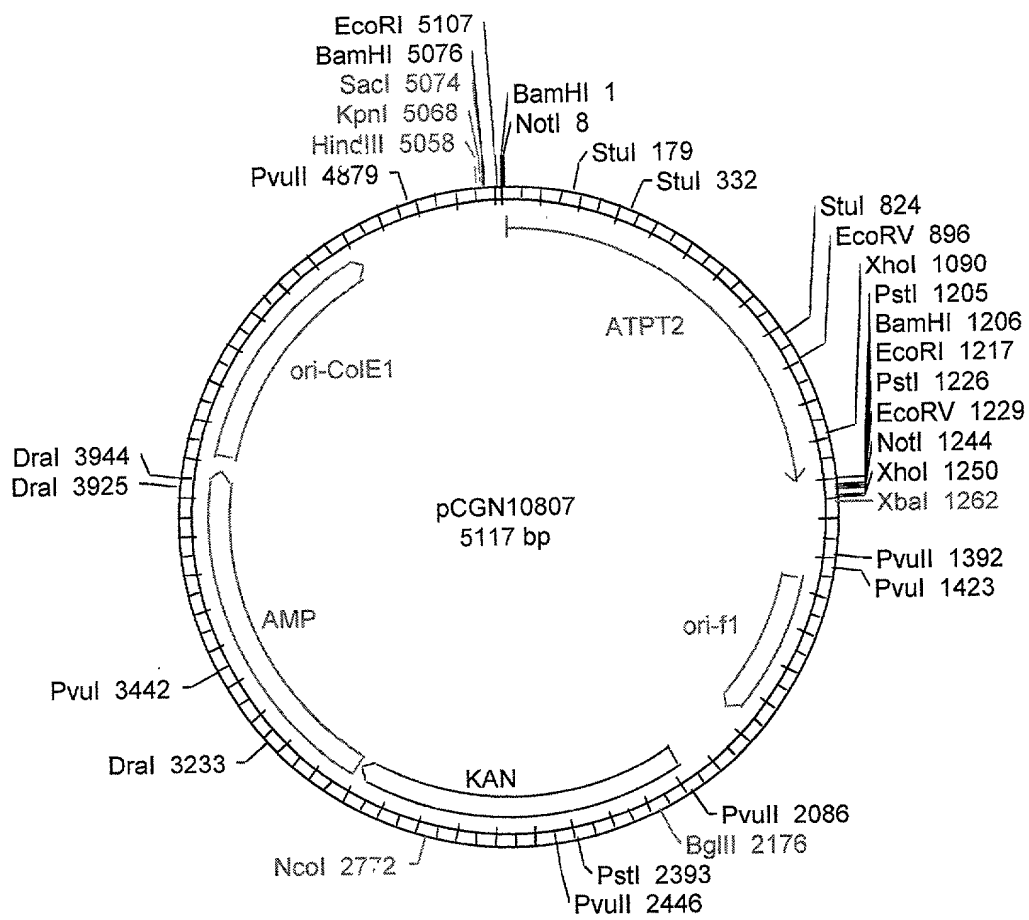


Figure 6

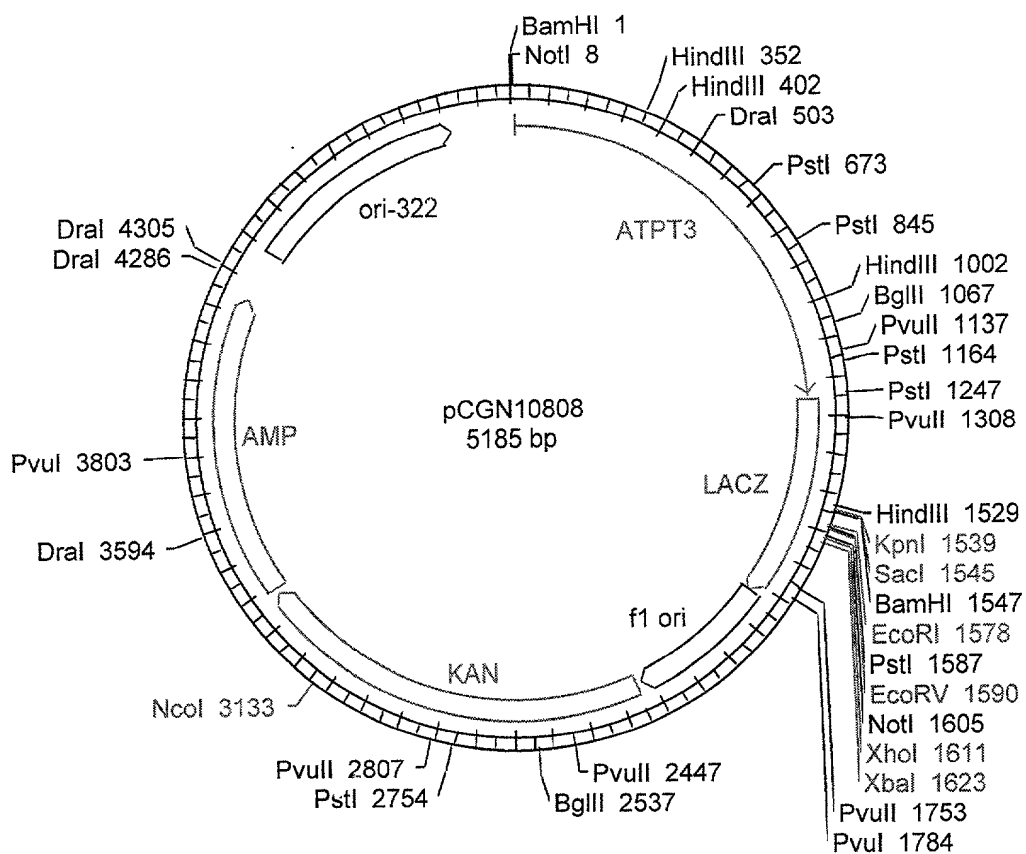


Figure 7

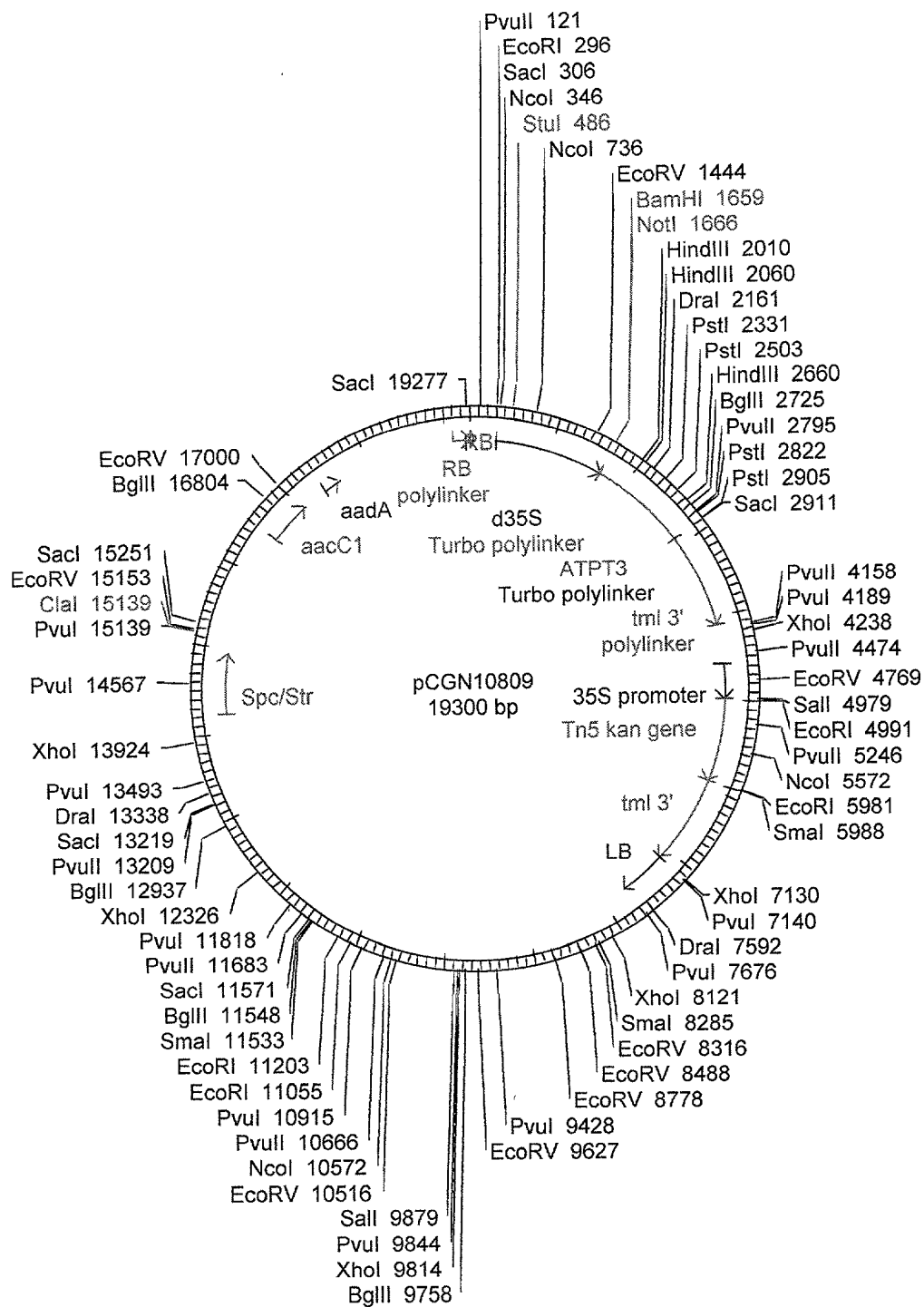


Figure 8

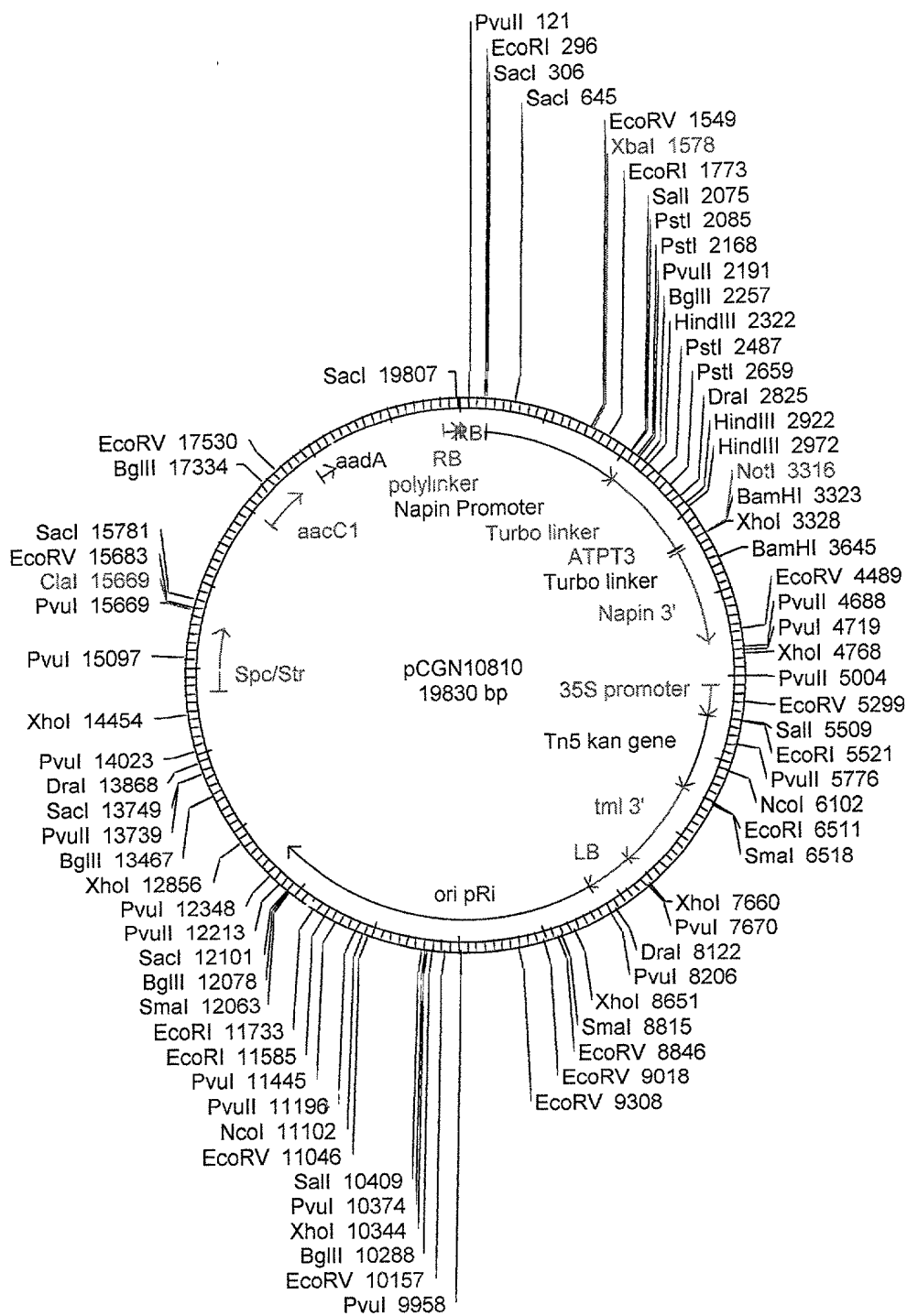


Figure 9

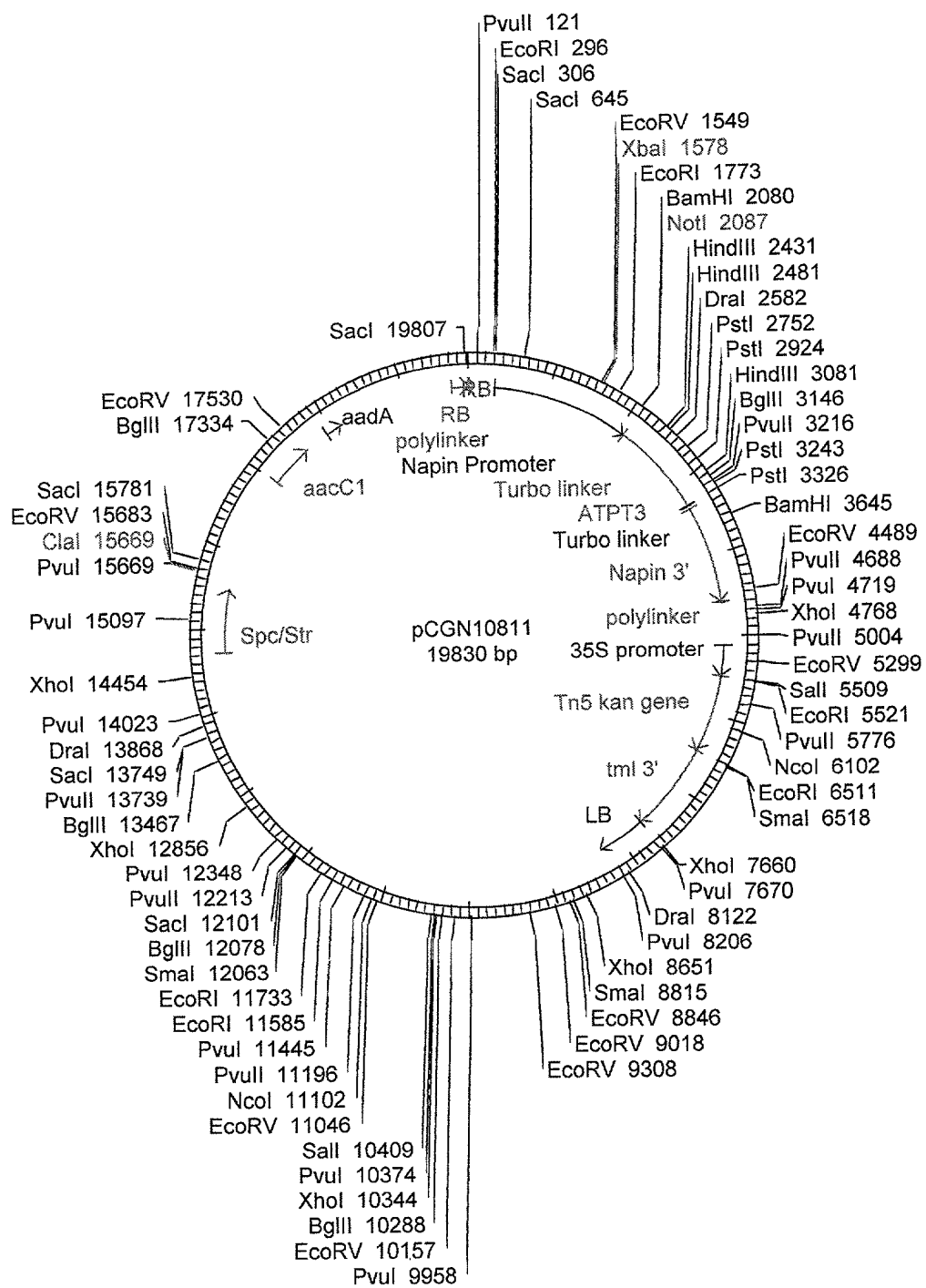


Figure 10

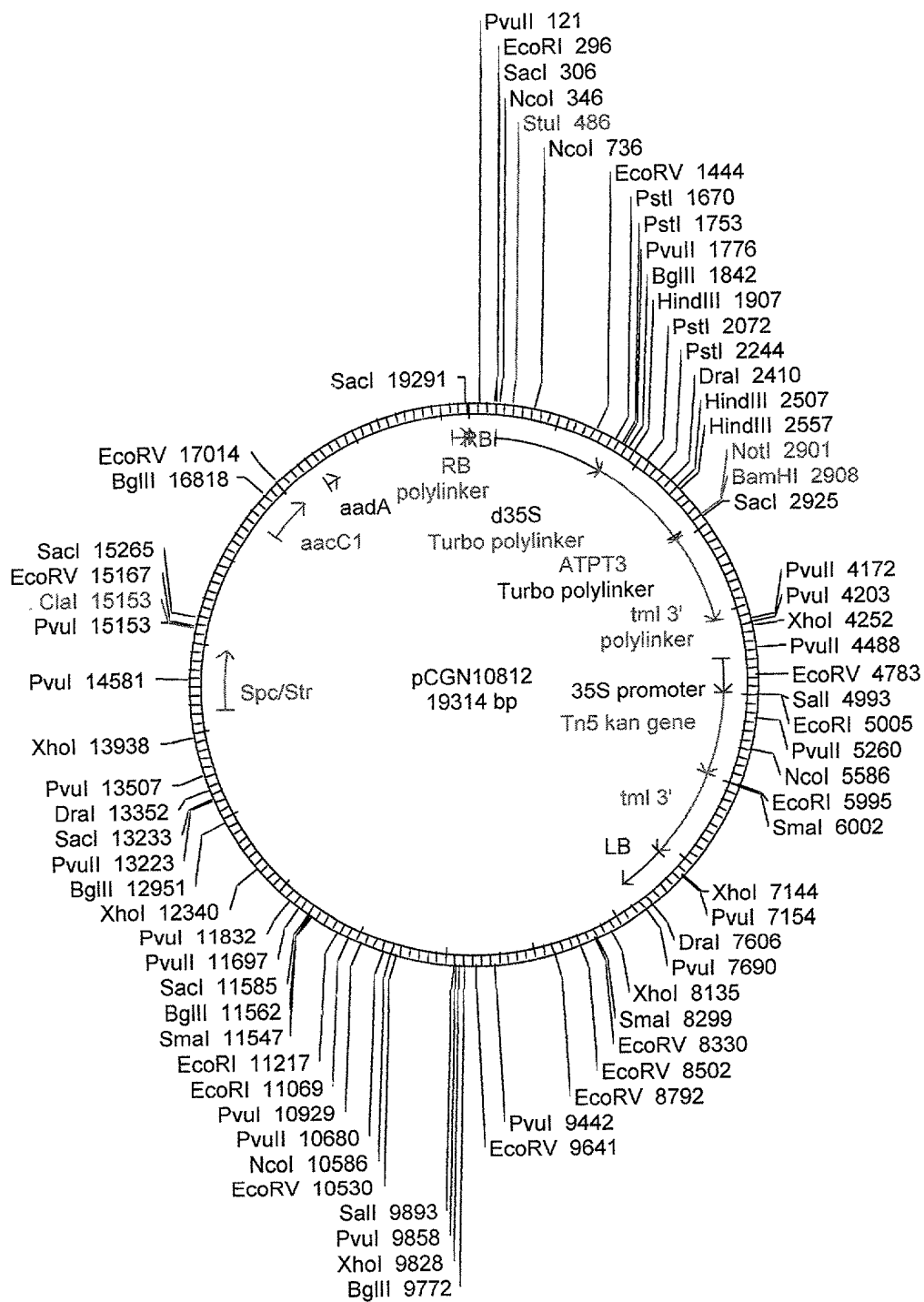


Figure 11

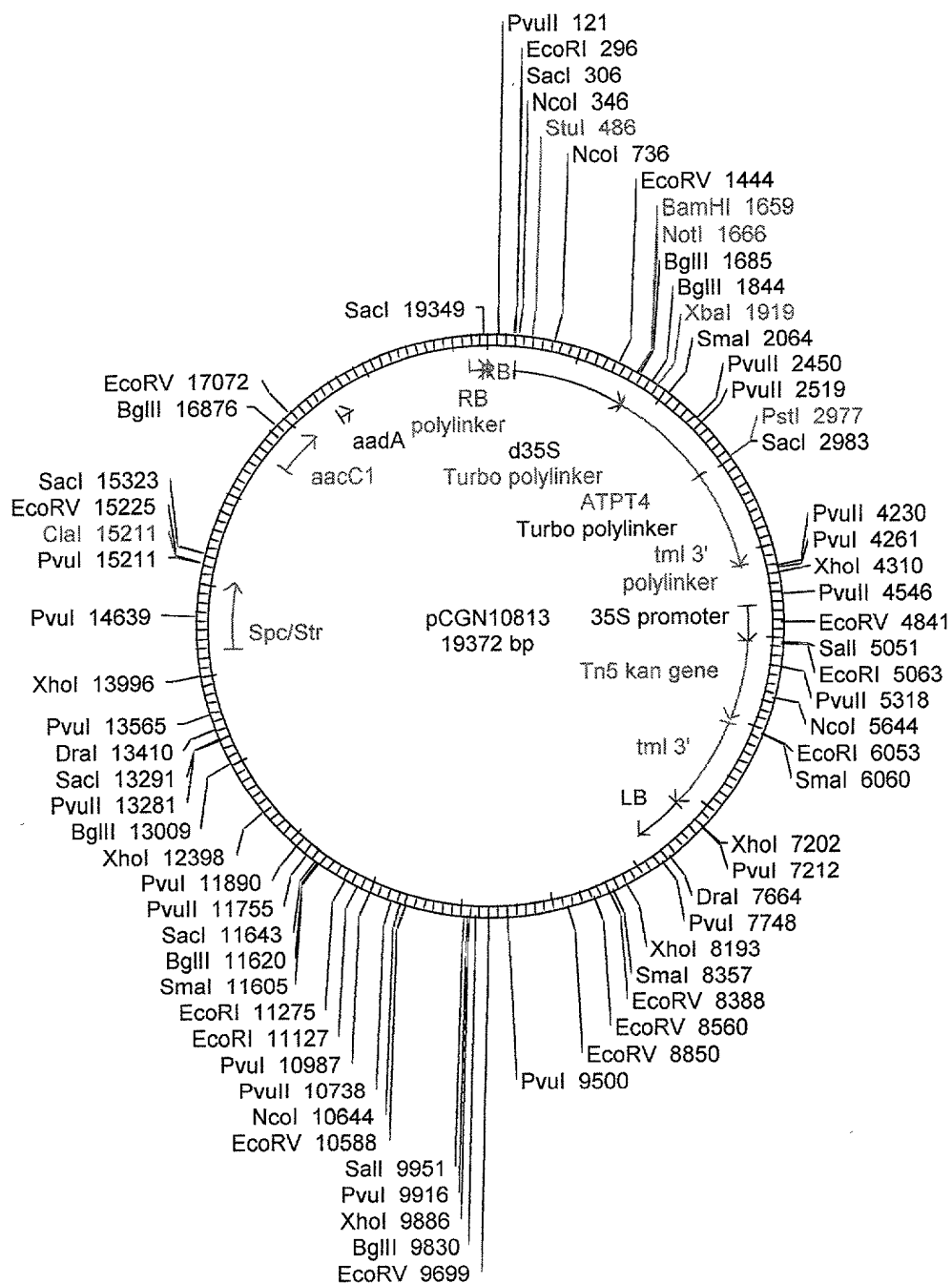


Figure 12

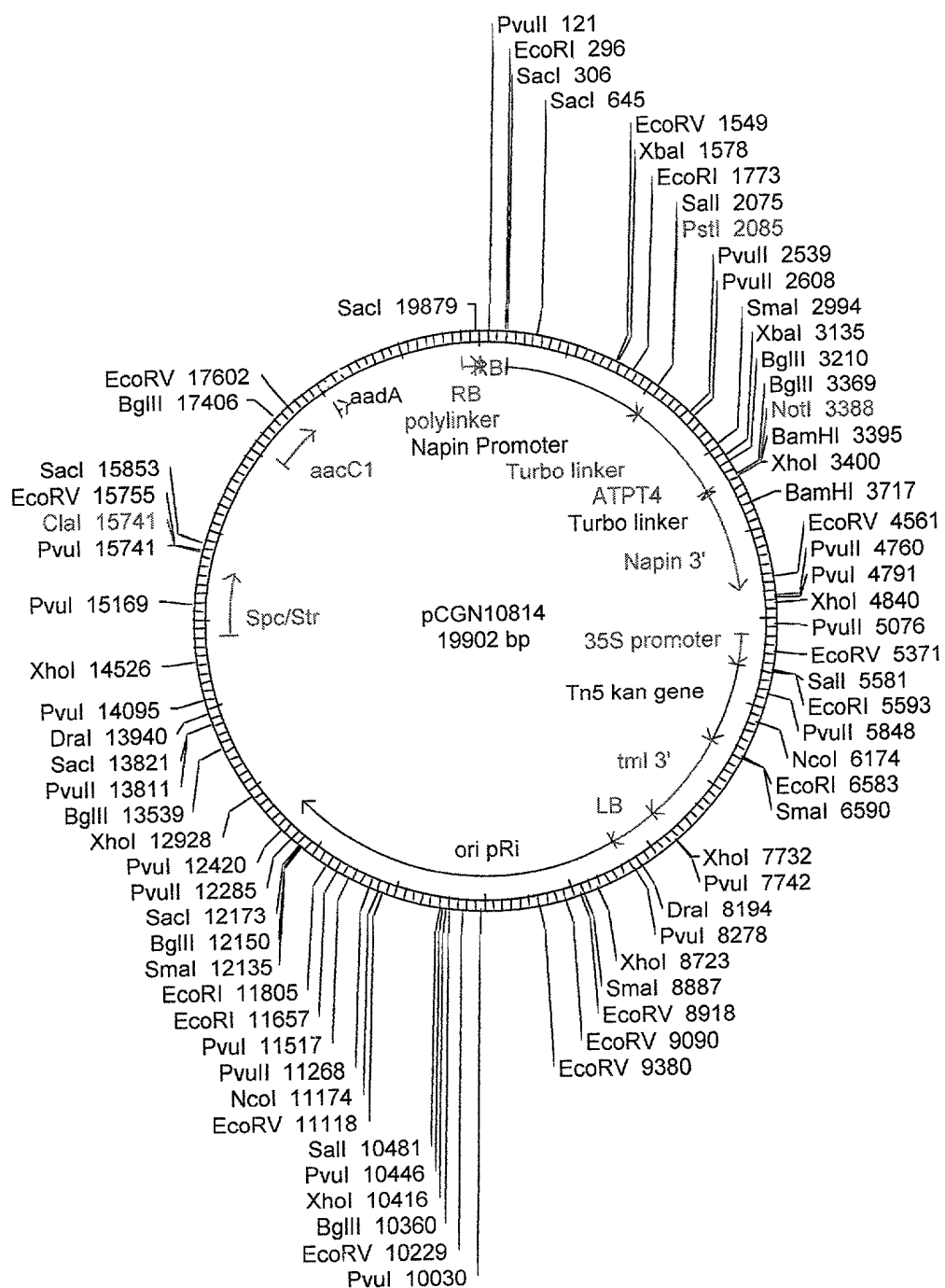


Figure 13

Figure 14

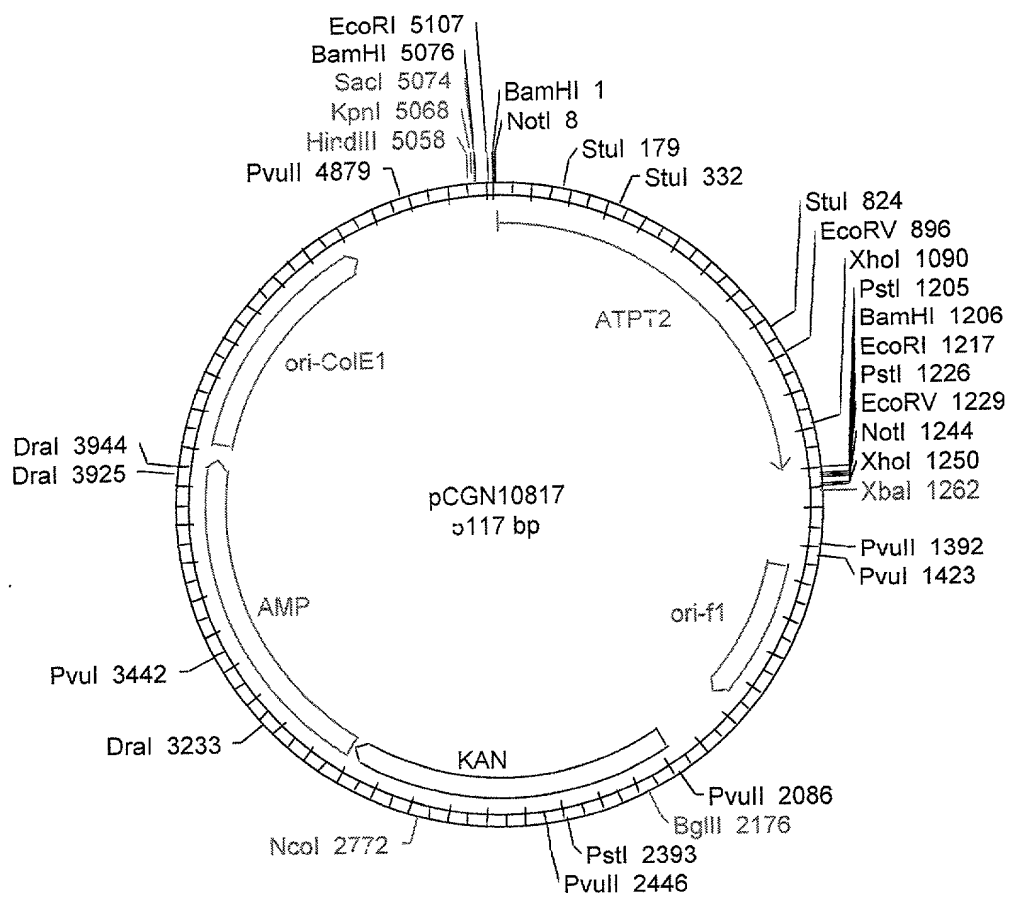


Figure 16

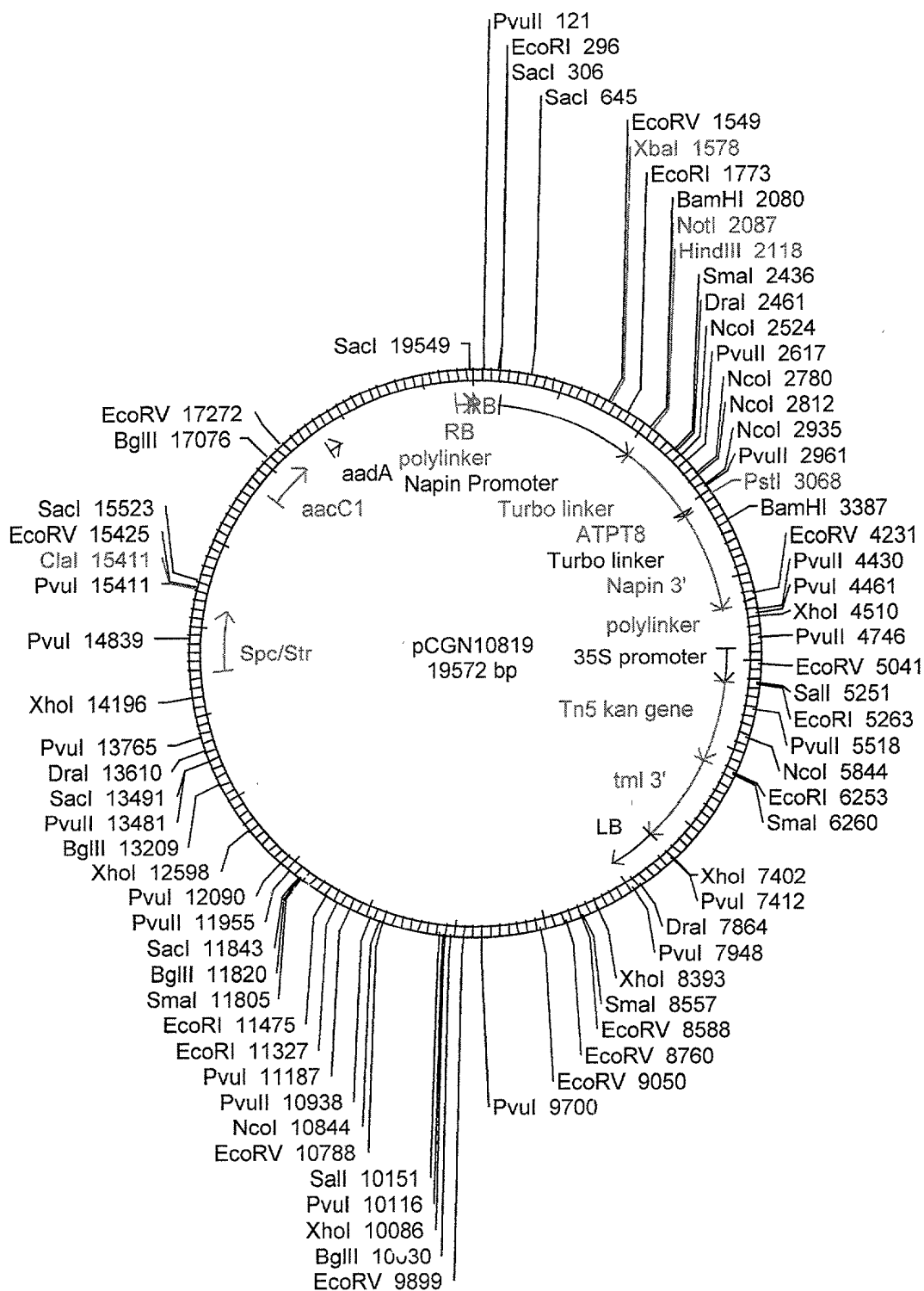


Figure 17

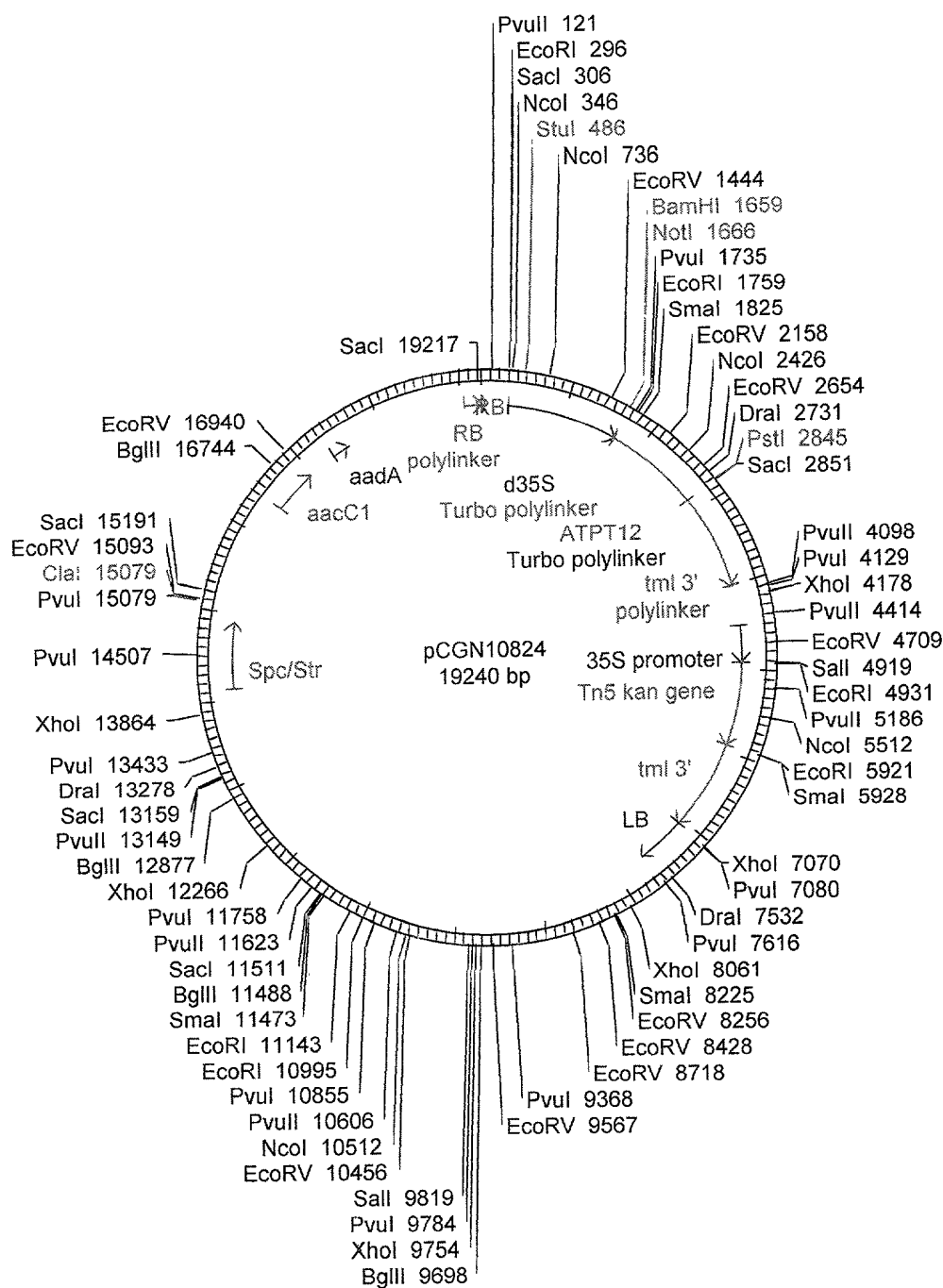


Figure 18

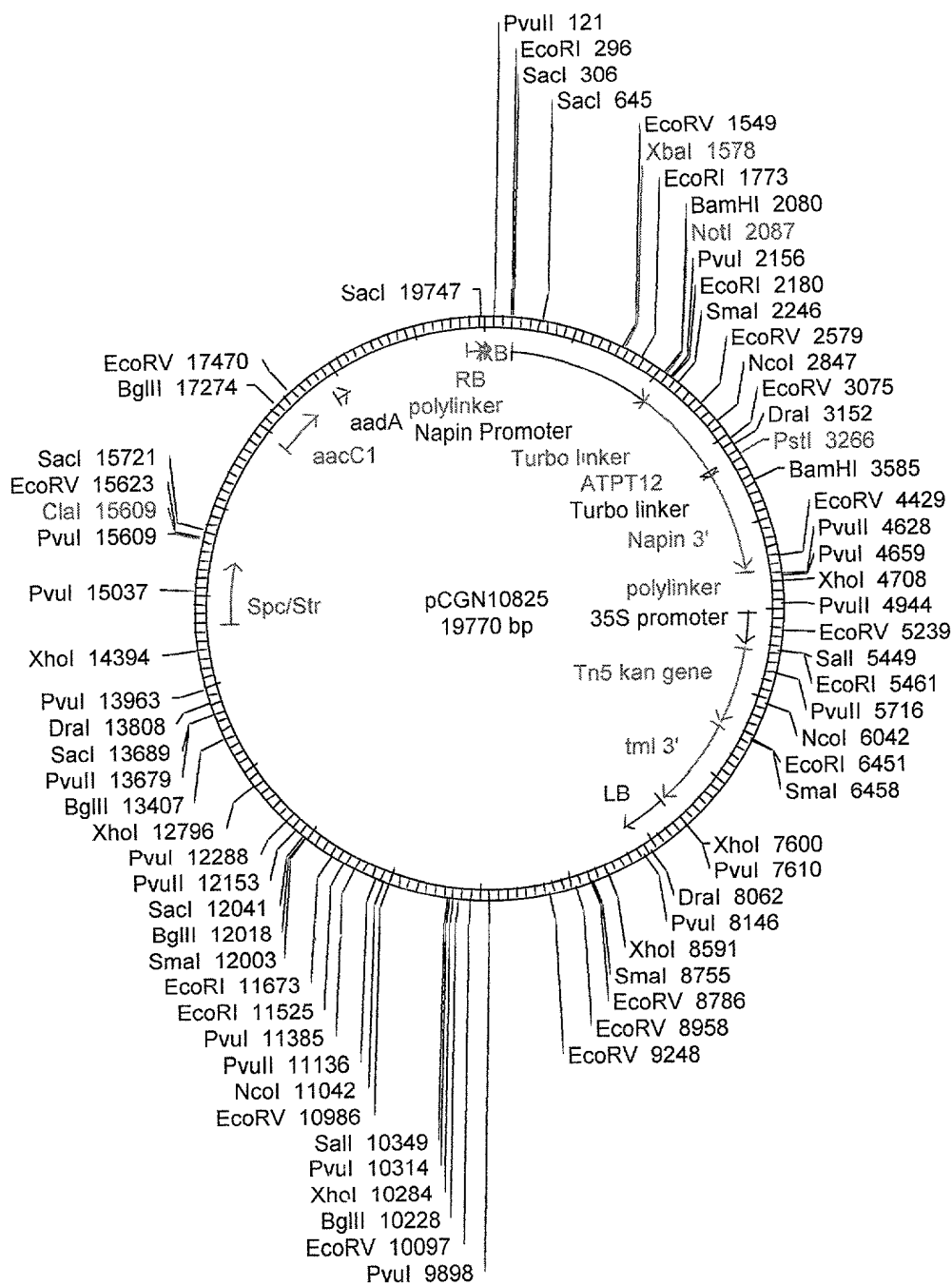


Figure 19

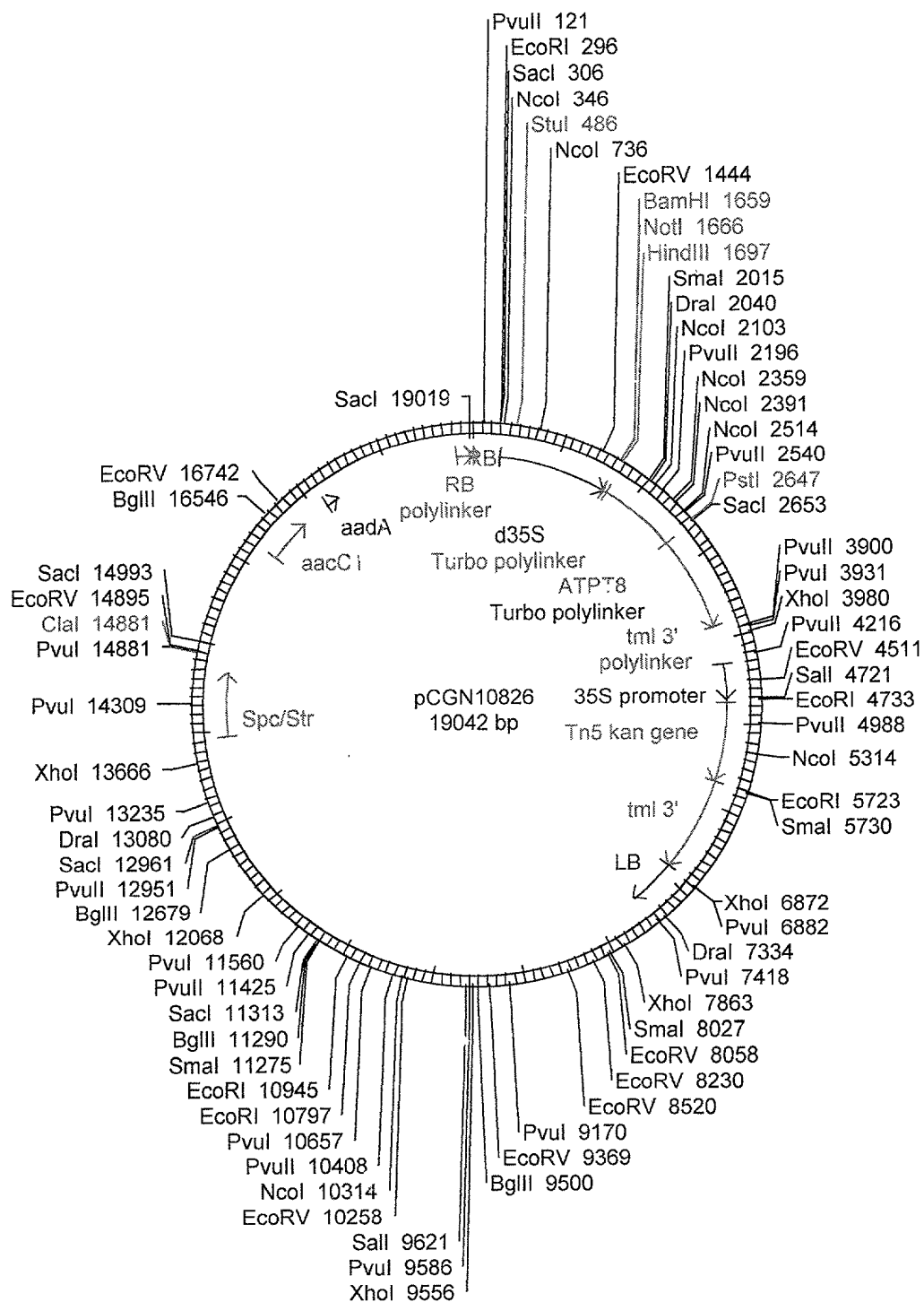


Figure 20


```

ATPT2 : -----*-----20-----*-----40-----*-----60-----*-----80
SLR1736 : -----MESLSSSVSAAGFCWKQNLKHLSELRLVLRCDSSKVAKPKFRNNLVRPDQSSSLLYPKHKSFRVNATAQ : 80
ATPT3 : -----MAFFGLSRVSRRLKSSSVSPDSSSSALLQSQHKSLSNPVTTYTNPTTKCPSWMDNQVWSKGRHLQEKFEFGWNYRLICGMSS : 89
SLR0926 : -----MMRKSVMYRFSRSISVSSSLPNPRLIPMSRELCAVNSFSQP-----PVSTESTAKLITGVRS DANVFATA : 67
ATPT4 : -----MMRKSVMYRFSRSISVSSSLPNPRLIPMSRELCAVNSFSQP-----PVSTESTAKLITGVRS DANVFATA : 67
SLR1899 : -----MTSINTVSTIHSRVTSDRVGVLRLNSDSVEFT-----RRSGFSTLIYSPGRFVRAEITD : 63
ATPT12 : -----MTSINTVSTIHSRVTSDRVGVLRLNSDSVEFT-----RRSGFSTLIYSPGRFVRAEITD : 63
SLR0056 : -----MSDT : 4
ATPT8 : -----MVA : 4
SLR1518 : -----MTES : 4

```

```

ATPT2 : -----*-----100-----*-----120-----*-----140-----*-----160-----*-----1
SLR1736 : -----PEAFDSNSKQ-----SFRDSDAFYR-----FSRPHITGTVLSLS-----VSFLAVEKVS-----DISPLFTGLE : 140
ATPT3 : -----SSVLECKPKDKDKSDGVVVKKASWDLVLEPEVRGYAKLARDKPIGTWILAWPCMS-----IALADPGS-----LPSEKYMALFG : 49
SLR0926 : -----MAAQTSSP-----PLWITLIL-----LRWHKPAGRIMIPALMA-----VCLAQ-----G-----LPPLPLGTAL : 170
ATPT4 : -----TAAATATATG-----EISSRVAALAGGHYAR-----CYWELSKARIMLVATSG-----TGIIIGTQNAISFPGCYTCA : 138
SLR1899 : -----TKIHRQHDSMG-----AVCKSYQULTK-----RIPLITTAASWIL-----ASEGR-----VDLPKLITLIG : 60
ATPT12 : -----DKVKSQTPDKAP-----AGSSINQILGKAS-----QETNWKIRQUTKPVTPPLVWGVCGAASGNFWTPEDVAKSILC : 139
SLR0056 : -----QNT-GQNOAKA-----ROLGKGAAP-----GESSIWKIRQMKPITWIPLIWGVCGAASSGCIWVSEDEFKALTC : 73
ATPT8 : -----EVPKLASAEY-----FFKRGQKQF-----RSTILLMATALNVRP-----EALIGEST-----DIVTSELVROR : 63
SLR1518 : -----SPLAPSTAPAT-----RKLMAAIKE-----PMYTVAVVPITVG-----SAVAYGLTG-----QWHDVFTIELL : 59

```

```

ATPT2 : -----80-----*-----200-----*-----220-----*-----240-----*-----260
SLR1736 : -----AVPALMNNIYVGQSSVEKVNKPILPLASGEYSVNTGIAVAFSFMFWLGMIVGSPFLWALFSEWGTAYS-INIPILR : 228
ATPT3 : -----AWACLLGNVYIVGQMVDRINKENILPLANGDEFSIAOGRMIVGLOGASTAIAMGLG-LMIGLTVGSLITGTAY---SVPEVR : 134
SLR0926 : -----GAL---IRGAGCTDLQD TKVDPRKURPIASGIAT-PEOGTGLGQILGIG-----ILQINNSRYLGAS-----SLHVF : 246
ATPT4 : -----GTAL---TSGIGCVDWLRDPÖVERTKORPLARALS-VQVGIGVALALLCAAG-----LAFYLPSEWICVA---AYPIV : 132
ATPT12 : -----TMIL---AASANSQFEISN SKMKRIMLRPLBSGRISVPHAVMATIAGASGACL-----LASKTMMAGLASAN---LVLIAF : 215
SLR1899 : -----GTAL---AASAQTCYQDYEMEKITARPIPAGKVÖPRHALAFALA GVISFAL-----LATFVNVSGLALSGL---IYFYL : 137
ATPT12 : -----MMSGPCTGYTQTDWYRD AINEPYRPIPSGAISEPEVITQWVLLGGEGIAAGILD-VWAGHTTPTVEYALG-----GSLISY : 223
SLR0056 : -----MLSGPIMTGYTQTDWYRD AINEPYRPIPSGAISEPEVITQWVLLGGEGIAAGILD-VWAGHTTPTVEYALG-----GSLISY : 223
ATPT8 : -----GIAE---ITEMIHVASLHMDVLDADITBRGVGSINVMGNKMSVAGDFTSRACGAL-----AALKNTEVALATAVEHLVTGETM : 157
SLR1518 : -----SALA---IIAMINLSDESDTIGIDVRKAHSVNIITGNRNVLFTLSNFTIAGYIGLMSMS---WRAQDWTLEITIGVA-----ITFYG : 144

```

Figure 22 1/2

Figure 22 1/2

```

*      280      *      300      *      320      *      340      *
ATPT2 : WKR-FALVAMCILA-RAIVQLAFLYH-IQTHVGRPT-ETRPL-ETAFMSFS-VIALFKDIPDIEGDKI-F-TR-ES-TL-QKR : 313
SLR1736 : LKR-FSILALCILT-RGIVNGLGFEF-FRIGLGYRPT-ITPIW-LTLFILVET-AALFKDVPDMEGDRQ-FKQ-LTQI-KON : 218
ATPT3 : SYP-LMKRFTFWPOAFLGLTINWGLG-WT-AVKGS-APSIY-PLYLSGVCMT-YYDTIYAHOKREDYK-VYKTA-RE-DNT : 328
SLR0926 : AYP-GAKRVFPVPO-LSIANGFAY-WS-AVTGD-UDATW-IMGATVFT-GETDVYAMADREDDR-I-VN-SAFF-QYV : 213
ATPT4 : VYT-PLKOLHPINTW-GAVVGAIPE-NG-WA-AASQO-GYNSK-IPAAIYFNO-PHFMALALCRNDYAA-GYKMS-FDP--S : 294
SLI1899 : VYTHWLKRHTAQNIV-GAAGSIPB-NG-WA-AVIGD-SWTPW-LFALIFLWPHFWALAIMIKDYAO-VNYMELP-IA-EKET : 220
ATPT12 : IYS-APPLKUKONGW-GNFAEGASY-SLPWAGQAHFST-PPDYV-LTLLYSIAG-GIAYVNDEKSVEGDRA-L-LO-LP-AF-TET : 308
SLR0056 : IYS-APPLKUKONGW-GNYALGASY-ALPWAGHALFT-NTPIW-LTLLYSIAG-GIAYVNDEKSVEGDRO-LEK-PP-MF-IGT : 242
ATPT8 : EITSTEQRYSMDY-OKTYKKTAS-ESNSCKAVAVLTGQFAEVA-LAFEYGRNLG-AFQILDDILDTGTSASE-KG-ESDIRH-GV : 231
SLR1518 : TYOGPFERLGLGLGELICLITFGP-AL-AAAYSSQSQSFNNET-PSVFGISATILFCSHFHQVEDDIA-A-CKK-PI-RL-TKL : 223
g      g

360      *      380      *      400      *      420      *      440
ATPT2 : VFWTCVTL-OMAYAVAILVGATSPFIMSKVISVGHVITATT-ARAKSVDSLSSKTEETS--CYMFIWKLFYAEYILPELK----- : 393
SLR1736 : VERGTLIL-TCGYLAMAINGWAMAMPNTAFLIVSHLCILAL-PWRSRDVHLESKTEIAS--FYOFIWKLFLELYLP-ALMLPNFS : 304
ATPT3 : KIMITGEGTASIGFIALSGFSADLGMWYASIAAASQOLGMO-CTADLSSGADCS--RKFSVSNKWFGALIFSG-VLGRSFQ- : 407
SLR0926 : GEAVGIFFALTIGCLFYBGMTIMNPLWLSIAFAI--VGWY-QYIQLSAPTEP-KTY-----GQIFGQNVITIGHVILHAG-LLGWL-- : 292
ATPT4 : GKRIAAVA-RNCFYMIPLGFIAYDWGLTSSWFCLESTLITLA-NATAFSFYRDRTHMA-----RKMFMASLFLPYFMSC-LHHRVSD : 379
SLI1899 : VSQIMWYS-LVPEFSLLIVPLHQIGILYLAIAIT--GGOF-VKAWOLKOAPGDRLA-----RGLKFKSIFYIMLICLA-VEDSLPV : 303
ATPT12 : AKMICVGA-DITQLSVAGILASGKPYALALVALI--IPOH-FOFKYFLKDPVKYDYK-----YQASAQPFVLGTFVTA-PASQH-- : 387
SLR0056 : AAMICVIM-DVFOAGIAGLYLVHQOLYATIVLELL--IPOITFQDMYFLRNPLENDYK-----YQASAQPFVLVFGMLATG-ALGHAGI- : 324
ATPT8 : ITAPILFA-EEFPOLREVVDQVEKDPNDVIALEYLGKSKGTQARELAMEHANLAALAGSLPETDNEDVKRSKRALID-THRVIITRN : 320
SLR1518 : GSQVLTLS-VSLYLITAIAGVICHQAPWQTLIL-IASLPWAVOL-RHYGVGYHDQPEQVSNCK--FIAVNLHFSGMLMA/GYGWAGLG-- : 307

```

```

*      460      *      480      *
ATPT2 : ----- :
SLR1736 : NTIF----- : 308
ATPT3 : ----- :
SLR0926 : ----- :
ATPT4 : NQOOLVEAGLTNSVSGEVKTQRKKRVAQPPVAYASAPFFFLPAPSFSYP : 431
SLI1899 : --HQVAQMGTLLG----- : 316
ATPT12 : ----- :
SLR0056 : ----- :
ATPT8 : K----- : 321
SLR1518 : ----- :

```

Figure 22 2/2

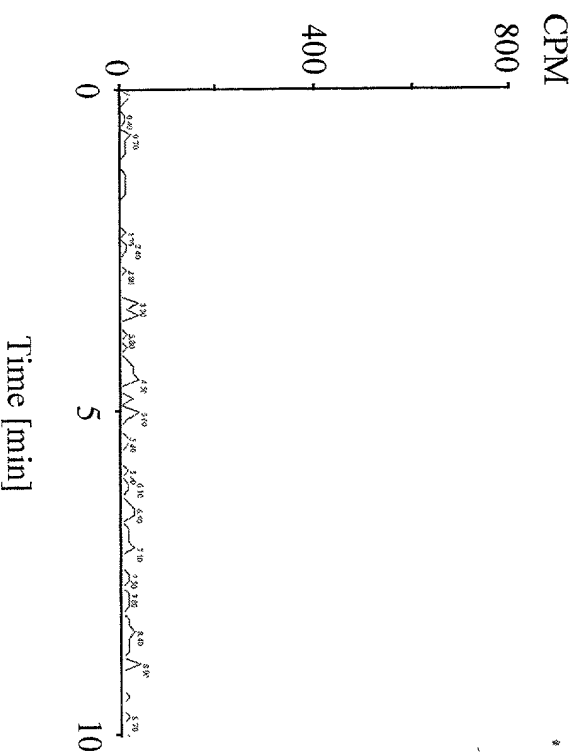
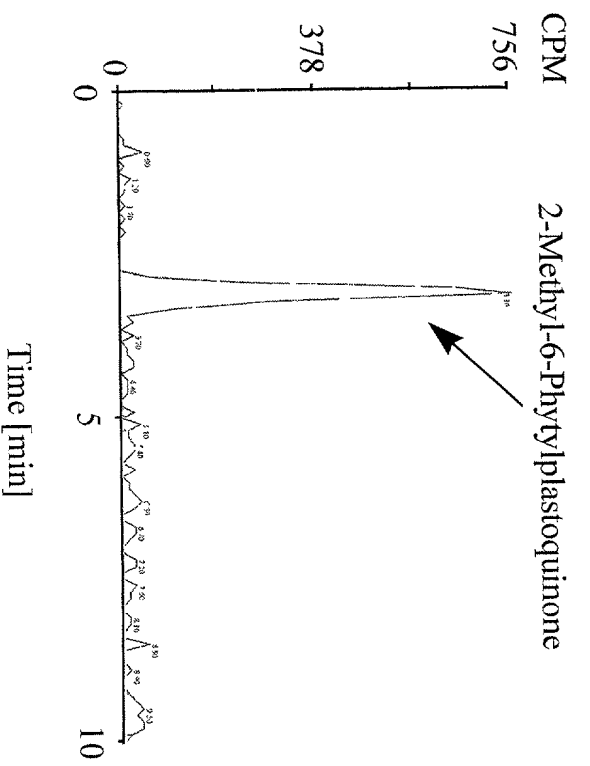


Figure 23

Synechocystis 6803 wild type

Synechocystis slr1736 knockout

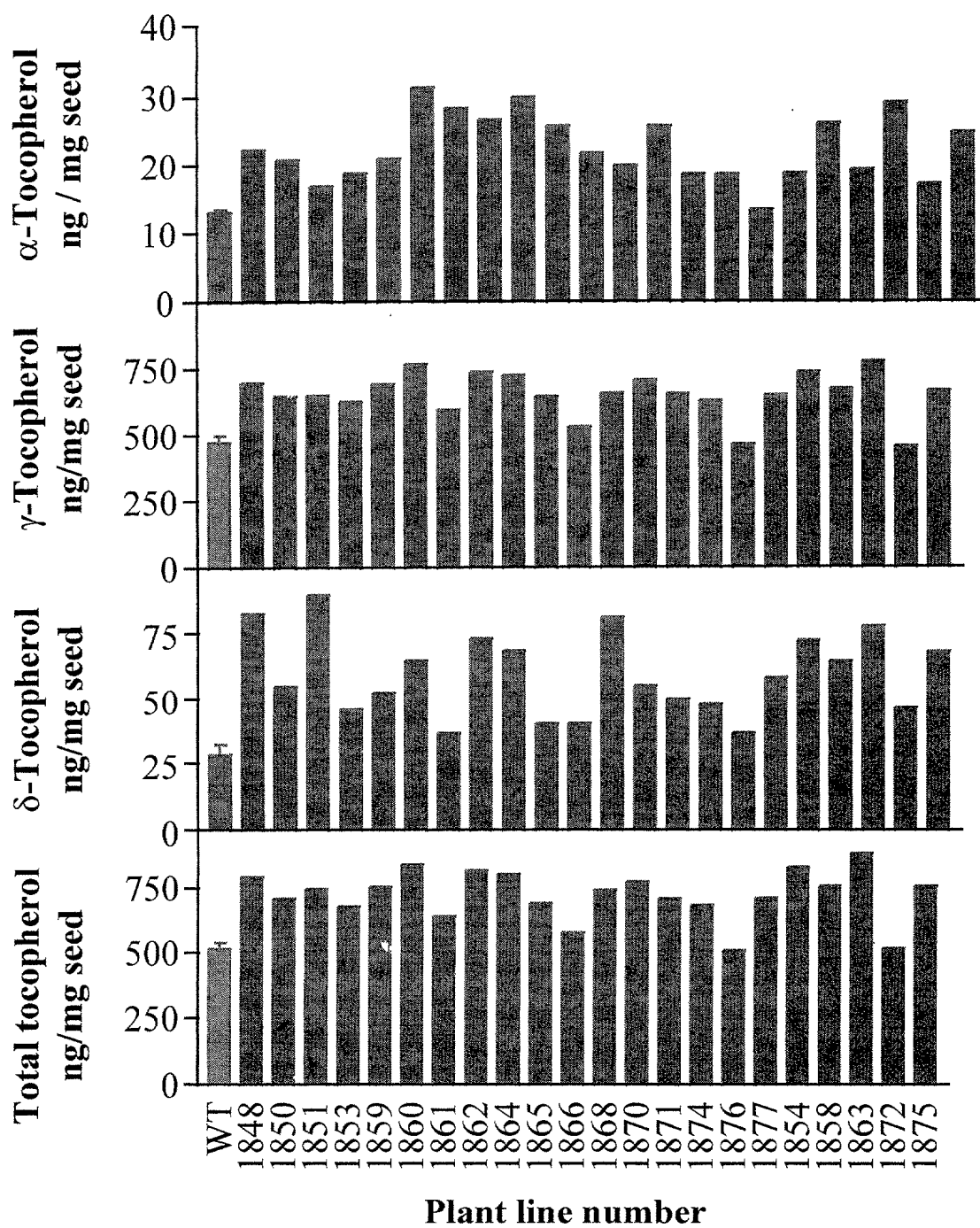


Figure 24

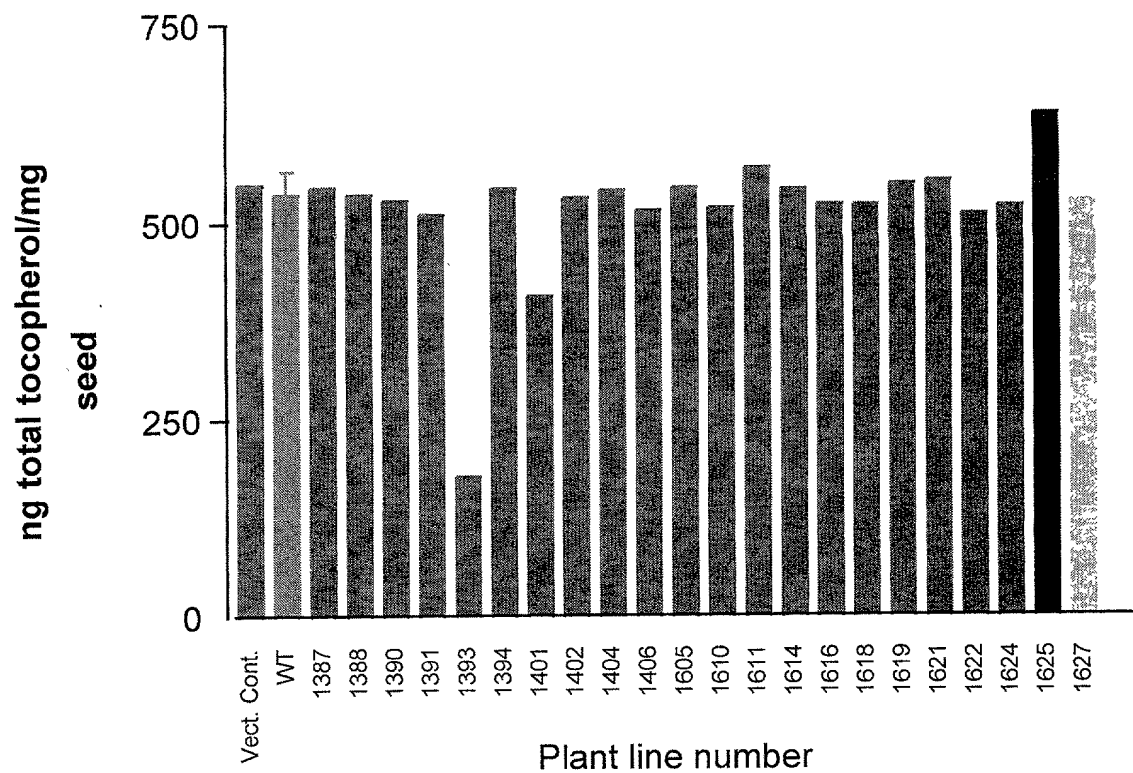


Figure 25

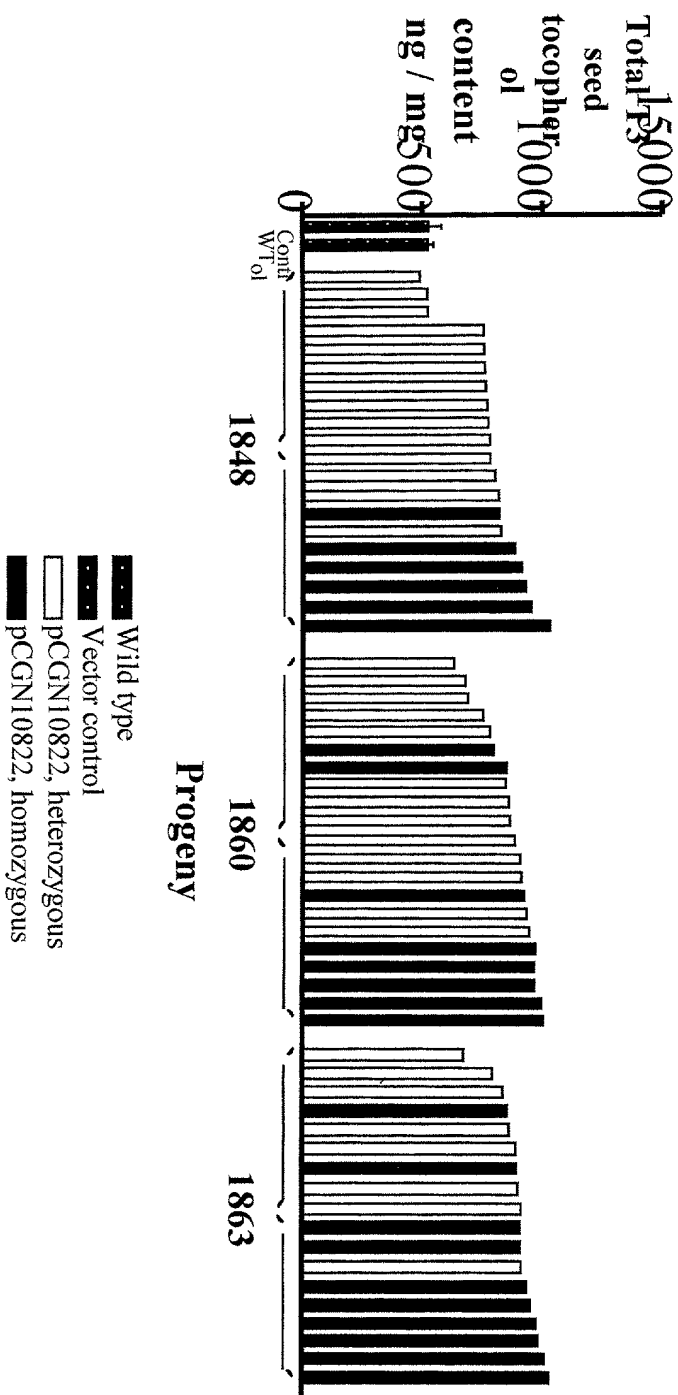


Figure 26

Total tocopherol in Napin ATPT2 Canola Seed

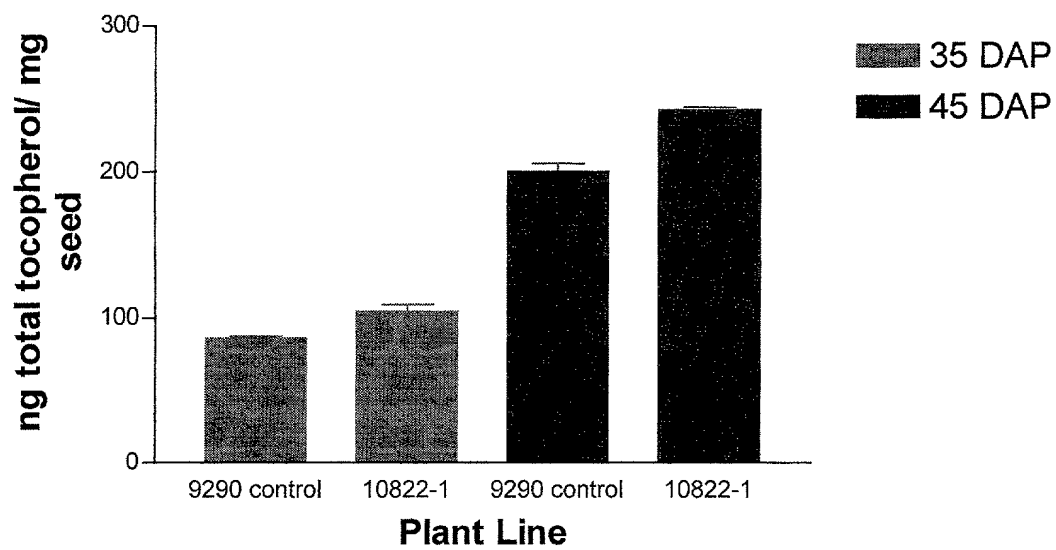


Figure 28

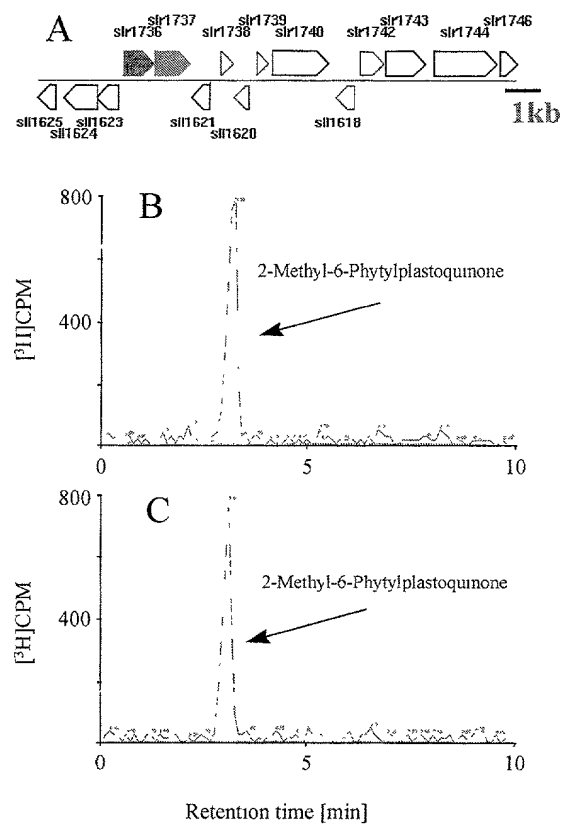


Figure 29

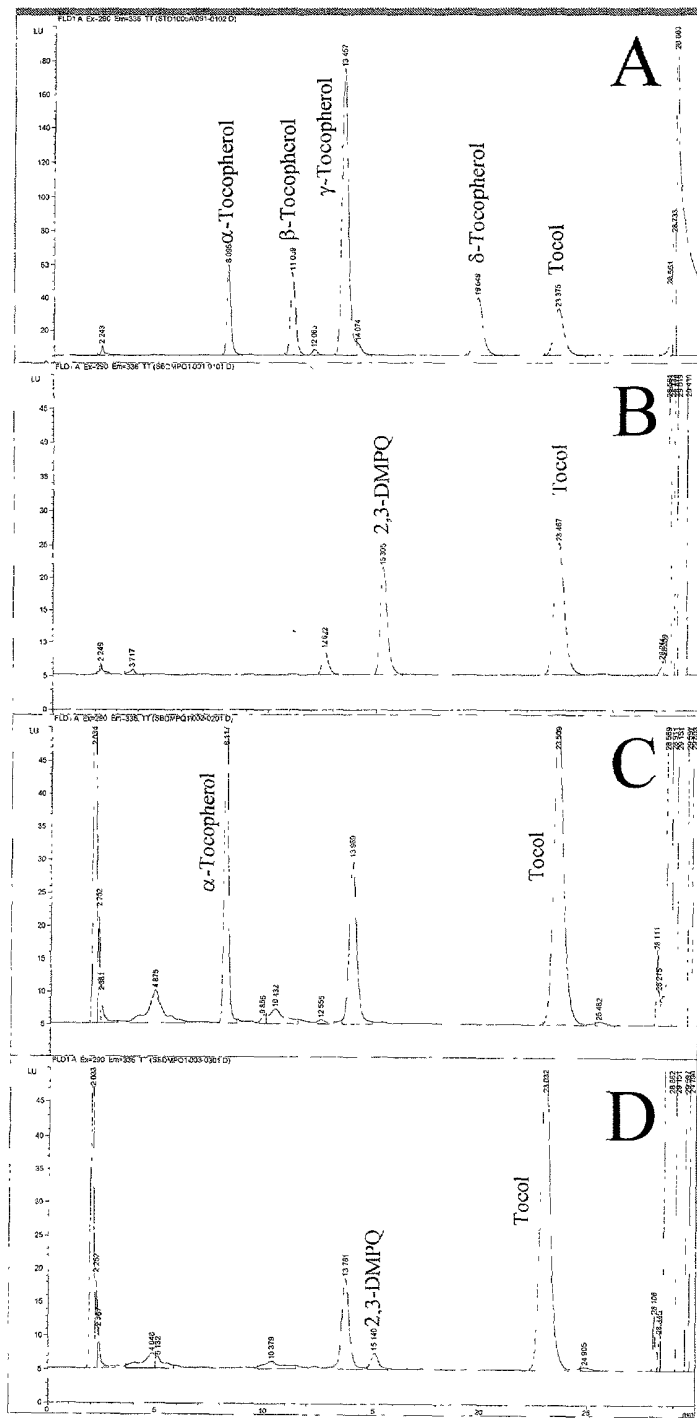


Figure 30

Variable	Mean	SD	Min	Max	Median	Q1	Q3	Mode	Skewness	Kurtosis	Jarque-Bera	Normality
Age	35.2	12.5	20	65	32	28	38	30	0.5	3.2	0.1	Normal
Gender	0.5	0.5	0	1	0.5	0.5	0.5	0.5	0.0	0.0	0.0	Normal
Education	12.5	2.5	9	16	12	11	13	12	0.2	1.5	0.0	Normal
Income	45000	15000	20000	80000	40000	35000	50000	40000	0.3	2.0	0.0	Normal
Health	0.8	0.2	0.5	1.0	0.8	0.7	0.9	0.8	0.1	0.5	0.0	Normal
Marital	0.7	0.4	0	1	0.7	0.6	0.8	0.7	0.2	1.0	0.0	Normal
Employment	0.9	0.1	0.5	1.0	0.9	0.8	1.0	0.9	0.1	0.5	0.0	Normal
Unemployment	0.1	0.3	0	1	0.1	0.1	0.2	0.1	0.3	1.5	0.0	Normal
Disability	0.05	0.2	0	1	0.05	0.05	0.1	0.05	0.5	2.0	0.0	Normal
Depression	0.2	0.4	0	1	0.2	0.1	0.3	0.2	0.4	1.0	0.0	Normal
Stress	0.3	0.5	0	1	0.3	0.2	0.4	0.3	0.3	1.0	0.0	Normal
Life Satisfaction	0.7	0.3	0.5	1.0	0.7	0.6	0.8	0.7	0.1	0.5	0.0	Normal
Healthcare Access	0.9	0.1	0.5	1.0	0.9	0.8	1.0	0.9	0.1	0.5	0.0	Normal
Healthcare Cost	15000	5000	10000	25000	14000	12000	16000	13000	0.2	1.5	0.0	Normal
Healthcare Quality	0.8	0.2	0.5	1.0	0.8	0.7	0.9	0.8	0.1	0.5	0.0	Normal
Healthcare Access	0.9	0.1	0.5	1.0	0.9	0.8	1.0	0.9	0.1	0.5	0.0	Normal
Healthcare Cost	15000	5000	10000	25000	14000	12000	16000	13000	0.2	1.5	0.0	Normal
Healthcare Quality	0.8	0.2	0.5	1.0	0.8	0.7	0.9	0.8	0.1	0.5	0.0	Normal

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
0	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99
0	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99

```

ATCEA4C371+ 362 AGAGAAGAGGGAGAGTTTTTTGTTTATGTATTCTGTGGAGAATCCTGCATTTCCGGCAGAG
PIR:T04448 46 E K R E S F C F M Y S V E N P A F R Q S

Query- 11415 TTTGTCACCATTGGAAGTGGCTCTATATGGACCTAGATTCACTGGTGTGGAGCTCAGAT
ATCEA4C371+ 422 TTTGTCACCATTGGAAGTGGCTCTATATGGACCTAGATTCACTGGTGTGGAGCTCAGAT
PIR:T04448 66 L S P L E V A L Y G P R F T G V G A Q I

Query- 11355 TCTTGGCGCTAATGATAAAATATTTATGCCAATACGAACAAGACTCTCACAATTTCTGGGG
ATCEA4C371+ 482 TCTTGGCGCTAATGATAAAATATTTATGCCAATACGAACAAGACTCTCACAATTTCT
PIR:T04448 86 L G A N D K Y L C Q Y E Q D S H N F W G
ATCEA4C371+ Exon 11538 11301 Confidence: 100 100

Query- 11295 AGGTAACCTCCTTGACCCCTTAAATGCTGTGTCATGACAATAAGAAATCATATCTGAGTCT
ATCEA4C371+ 537
PIR:T04448 106 D
PIR:T04448 Exon 11609 11294 Confidence: 100 100

Query- 11235 TTTCTCTACTTCTAGTACTAATGTTTCGTTATTGTTGTAAAGATCTAAGTCTTATCTGAA
PIR:T04448 107

Query- 11175 TTTTGTACATTTTGGTTCTGGTGTCTTCTCAACATGAATTTGTATATATGACTTTAAAG
PIR:T04448 107

Query- 11115 ATTGCTTACCTAAAGTTTTACTCATGCTAGATCGACATGAGCTAGTTTGGGGAATAC
PIR:T04448 107 R H E L V L G N T

Query- 11055 TTTTGTGCTGTGTGCCAGGCGCAAAGGCTCCAAACAAGGAGGTCCACCAGAGGTCTCTCAC
PIR:T04448 116 F S A V P G A K A P N K E V P P E
PIR:T04448 Exon 11083 11004 Confidence: 96 100

Query- 10995 TCCTCCCTTGTTGGTTACTTTGTTATCTGTTAAATAGTTTCCAATTGTATCCGGATAGT
PIR:T04448 133

Query- 10935 GTTCTACTTCTCCTTGTAGAAAATCTCAAGTTTTTGTACTCTTGCTATTCTCTTGGATG
PIR:T04448 133

Query- 10875 TTGATTTGTAAAGCATGTCGTTTTATTGTAGGAATTTAACAGAAGAGTGTCCGAAGGGTT
PIR:T04448 133 E F N R R V S E G F

Query- 10815 CCAAGCTACTCCATTTTGGCATCAAGGTCACATTGCGATGATGGCCGGTAATTATATGA
PIR:T04448 143 Q A T P F W H Q G H I C D D G R
PIR:T04448 Exon 10844 10768 Confidence: 100 100

Query- 10755 TTCTATGCACAACAAGAATTCACATATATTATAAATATTGGATATTGAGTATTTTGTGTA
PIR:T04448 159

Query- 10695 AAATTTCTGTGTTTAAATCTGACTTGACTTGTGTTGTCAGTACTGACTATGCGGAAACTG
PIR:T04448 159 T D Y A E T V

```

Figure 31
2/5

Query- 10635 TGAAATCTGCTCGTTGGGAGTATAGTACTCGTCCCGTTTACGGTTGGGGTGATGTTGGGG
 PIR:T04448 166 K S A R W E Y S T R P V Y G W G D V G A

Query- 10575 CCAACACAGAAGTCAACTGCAGGCTGGCCTGCAGCTTTTCCTGTATTTGAGCCTCATTGGC
 PIR:T04448 186 K Q K S T A G W P A A F P V F E P H W Q

Query- 10515 AGATATGCATGGCAGGAGGCCTTTCCACAGGTGTGAGCTTTGCTTGATTGACTTAAAGTT
 PIR:T04448 206 I C M A G G L S T G
 PIR:T04448 Exon 10655 10486 Confidence: 96 100

Query- 10455 AATAAATAGACGGTTAAGTTTACTTGCCTAGTACTAACAGAAAATTAGAAAGAAACCAC
 PIR:T04448 216

Query- 10395 CCTCTTTCTATCAGCAGAACTGCTATTGTAGTTCCTTATTTTCTCTGTATTGTCAGG
 PIR:T04448 216

Query- 10335 GTGGATAGAATGGGGCGGTGAAAGGTTTGAGTTTCGGGATGCACCTTCTTATTCAGAGAA
 PIR:T04448 216 W I R W G G E R F E F R D A P S Y S E K

Query- 10275 GAATTTGGGGTGGAGGCTTCCCAAGAAAATGGTTTGGGTTAAACATTTTCATCCTTTTGCT
 PIR:T04448 236 N W G G G F P R K W F W
 PIR:T04448 Exon 10336 10239 Confidence: 96 100

Query- 10215 ACATTTCTTGTGCAGACTTTAGTTAGCTAGTGGACCTGTGTATACACCCACATGTAGTA
 PIR:T04448 248

Query- 10155 TACTTGTGTTGATAGCTTTATTTGTCAATGTCTCTTTACAGGTCCAGTGTAATGTCTTTGA
 PIR:T04448 248 V Q C N V F E

Query- 10095 AGGGGCAACTGGAGAAGTTGCTTTAACCAGAGGTGGCGGGTTGAGGCAATTGCCTGGATT
 PIR:T04448 255 G A T G E V A L T A G G G L R Q L P G L

Query- 10035 GACTGAGACCTATGAAATGCTGCACTGGTATGCACTTATAAGATCTTCTTAAGCAATGA
 PIR:T04448 275 T E T Y E N A A L
 PIR:T04448 Exon 10115 10008 Confidence: 100 100

Query- 9975 CAGTGAGTATTAGAAGGCAGATAGTTTACAAAAGCTCTGGGCCCTTGTAATCTGCAGGT
 PIR:T04448 284 V

Query- 9915 TTGTGTACACTATGATGGAAAAATGTACGAGTTTGTTCCTTGGAAATGGTGTGTTAGATG
 PIR:T04448 285 C V H Y D G K M Y E F V P W N G V V R W
 GSDB:S:495- 532 tagatg

Query- 9855 GGAAATGTCTCCCTGGGG TTATTGGTATATAACTGCAGAGAACGAAAACCATGTGGTAA
 PIR:T04448 305 E M S P W G Y W Y I T A E N E N H V
 GSDB:S:495- 526 ggaaat tctccctgggggttattggtatataactgcagagaNcgNaaacctatgtg
 PIR:T04448 Exon 9917 9801 Confidence: 100 100
 GSDB:S:495- Exon 9861 9801 Confidence: 93 93

Query- 9796 ATTTGTTTACTAGTTTCATTTCAGTTTACTTTTGACATCATATCATTCCCTTATGGCTA

Figure 31
 3/5

```

PIR:T04448      323  -----
GSDB:S:495-    471  -----

Query-          9736  GATTCCAACACCCGATGAATGTCTTGTGACAGGTGGAAGCTAGAGGCAAGAACAATGAAG
PIR:T04448      323  V E L E A R T N E A
GSDB:S:495-    471  gtggaactagaggcNagaacaaatgaag

Query-          9676  CGGGTACACCTCTGCGTGTCTCTACCACAGAAGTTGGGCTAGCTACGGCTTGCAGAGATA
PIR:T04448      333  G T P L R A P T T E V G L A T A C R D S
GSDB:S:495-    443  cgggtacacacctgctgctgctcctaccacagaagttgggctagctacggcttgcagagata

Query-          9616  GTTGTTACGGTGAAATTGAAGTTGCAGATATGGGAACGGCTATATGATGGAAGTAAAGGCA
PIR:T04448      353  C Y G E L K L Q I W E R L Y D G S K G K
GSDB:S:495-    383  gttgttacggtgaattgaagttgcagatatgggaacggctatatgatggaagtaaaggca

Query-          9556  AGGTATGTATGCTAATGTGATCCAATCCCTGTAGTTAAAGTCTTAAACAAATCCTAAGGC
PIR:T04448      373  L K V L T N P K A
GSDB:S:495-    323  ag
PIR:T04448      Exon    9704    9555  Confidence: 100 100
GSDB:S:495-      Exon    9704    9555  Confidence: 98 100

Query-          9496  AGTGAAAGAAGATTATGAACGTTTGGTTATGGTTAAACAATGATGCAGGTGATATTAGAGAC
PIR:T04448      382  V K E D Y E R L L W L T M M Q V I L E T
GSDB:S:495-    321  gtgatattagagac

Query-          9436  AAAGAGCTCAATGGCAGCAGTGGAGATAGGAGGAGGACCGTGGTTGGGACATGGAAAGG
PIR:T04448      402  K S S M A A V E I G G G P W F G T W K G
GSDB:S:495-    307  aaagagctcaatggcancagtgagataggaggaggaccgtgggttgggacatggaaagg

Query-          9376  AGATACGAGCAACACGCCCCGAGCTACTAAACAGGCTCTTCAGSTCCCATTTGGATCTTGA
PIR:T04448      422  D T S N T P E L L K Q A L Q V P L D L E
GSDB:S:495-    247  agatacgagcaacacgcccagctactaaaacaggctcttcagggtccattggatcttga

Query-          9316  AAGCGCCTTAGGTTTGGTCCTTTCTTCAAGCCACCGGGTCTG TAA
PIR:T04448      442  S A L G L V P F F K P P G L
GSDB:S:495-    187  aagcgccttaggttttggtcccttcttcaagccaccgggtctgtaacattgatgagtggt
PIR:T04448      Exon    9522    9274  Confidence: 100 100

Query-          9256  TTTTCTGCTGATAGAGCCCATCTGATGAATGAGGCTTACTGATCTCATCTCTACCTT
PIR:T04448      456  -----
GSDB:S:495-    127  ttgtttgttgatagagaccatgtgatgaatgaagccttagtcatgtcattgctagcttc

Query-          9196  ACTATTATGTATGTATGATTTTAGTTCGTTCCGTCCTTGTGGTAAATGATACGGGCCAGT
GSDB:S:495-      67  actattatgtatgtatgatttttagttcgttcggtccttgtggtaaatgatacgggccagt

Query-          9136  GTAAAGTCTAGTTCAATAAAAGCCTTGAGTCGCATAATTTCAATTTCAAATTGCATC
GSDB:S:495-      7  gtaaaagt
GSDB:S:495-      Exon    9450    9130  Confidence: 98 100

```

Figure 31

ATCEA4C37145_1 3063693/emb|CAA18584.1| 4.0e-43 (AL022537) putative protein
[Arabidopsis thaliana]

PIR:T04448 sPIR-T04448 shypothetical protein F4D11.30 - Arabidopsis thaliana;
g3063693|emb|CAA18584.1 (AL022537) putative protein [Arabidopsis thaliana]_F4D11.30

GSDB:S:4955486|AI995392|AI995392|701673779 A. thaliana, Columbia Col-0, inflorescence-
1 Arabidopsis thaliana cDNA clone 701673779, mRNA sequence.

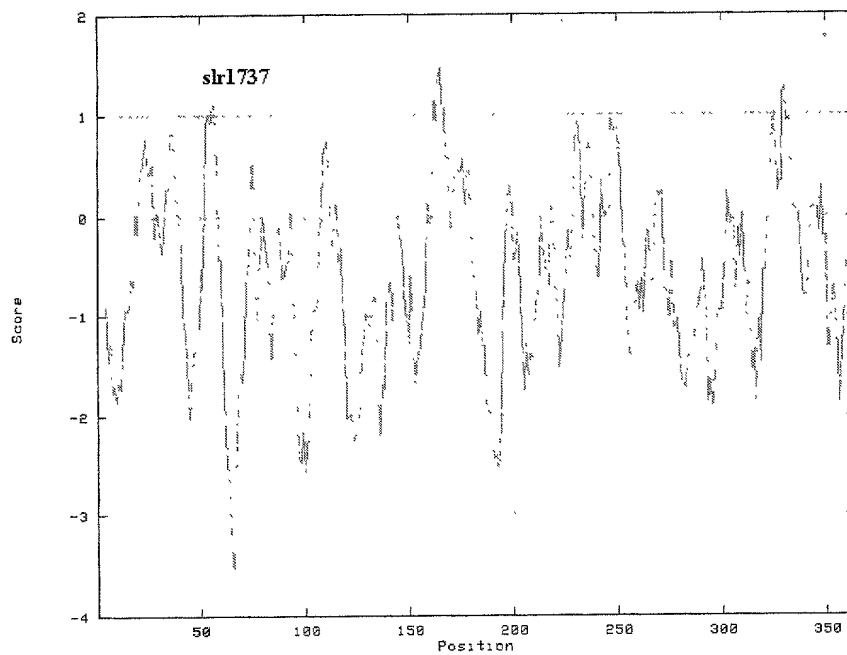


Figure 32

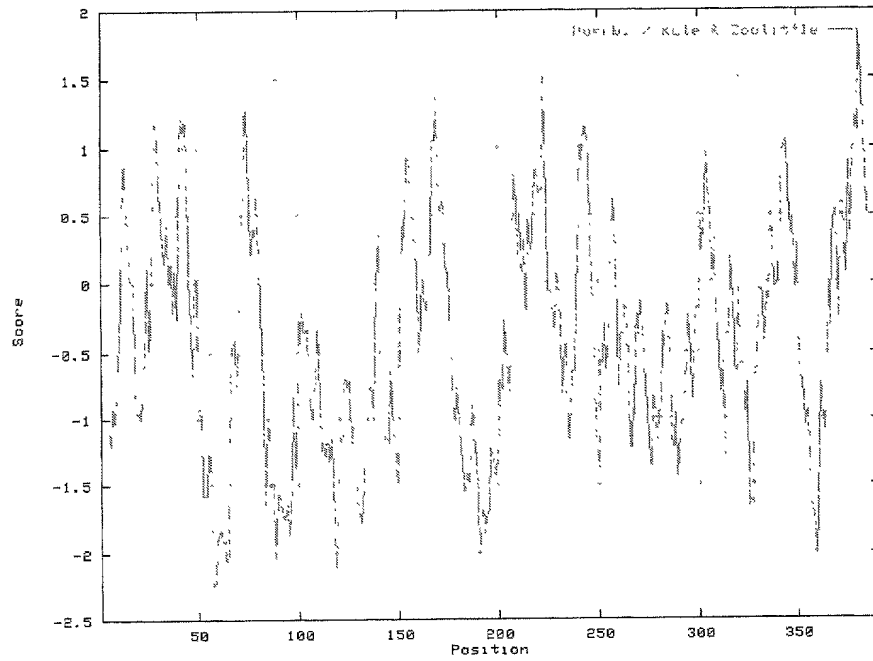


Figure 33

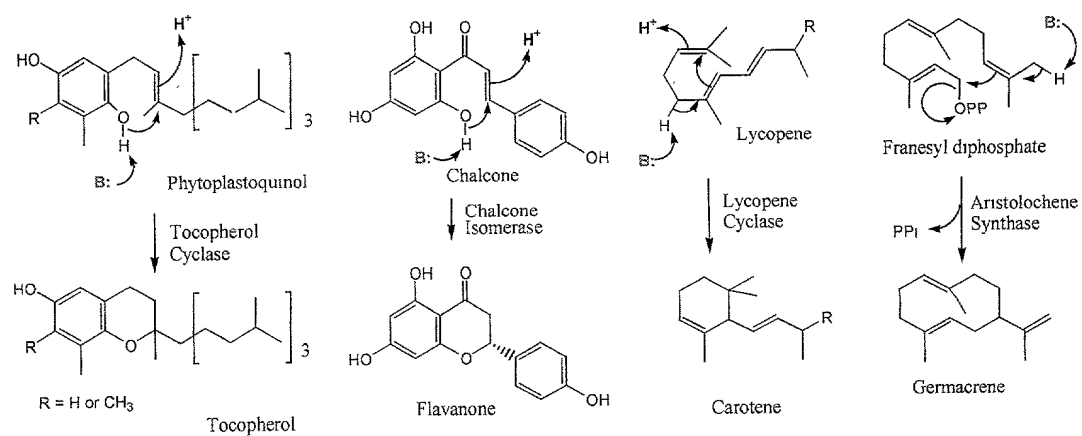


Figure 34

slr1737_SYNSP_S74814_ slr1737_ARATH_T04448_ CFI_ARATH_P41088_	-----M MEIRSLIVSMNPNLSSFELSRPVSPLTRSLVPRSTKLVPRSISRVSASI -----
slr1737_SYNSP_S74814_ slr1737_ARATH_T04448_ CFI_ARATH_P41088_	KFP-----PHSGYHWQQQS-PFFEGWYVRL STPNSETDKISVKPVYVPTSPNRELRTPHSGYHFDGTERKFFEGWYFRVS -----
slr1737_SYNSP_S74814_ slr1737_ARATH_T04448_ CFI_ARATH_P41088_	LPQSGESFAFMYSIENPASDHYGGGAVQILGPATK----KQENQEDQLV IPEKRESFCFMYSVENPAFRQSLSPLEVALYGFRTGVGAQILGANDKYL MSSSNACASPSPPA----VTKLHVDSV-
slr1737_SYNSP_S74814_ slr1737_ARATH_T04448_ CFI_ARATH_P41088_	WRTFPSVKKFWASPRQFALG-HWGKCRDNRQ-AXPLLSEEFFATVKEGYQ CQYEQDSHNFWGDRHELVTGNTFSAPVPAKAPNKEVPPEEFNRRRVSEGFQ --TFVPSVKSPASSNPFLG-GAGVRGLDIQ-GX-----IVIFTVIGVY
slr1737_SYNSP_S74814_ slr1737_ARATH_T04448_ CFI_ARATH_P41088_	IHQNHQGOI IHGDR-----HCRWQFTVEPEVTWGSNRPFRATAGW ATPFWHQGHICDDGRDYAETVKSARWEYSTRPVYWGVDVGAKQKSTAGW LEGNAVPSLSV-----KWKGKTTEELTESIPFFREIVIGAF
slr1737_SYNSP_S74814_ slr1737_ARATH_T04448_ CFI_ARATH_P41088_	LSFLPLFDPGWQIILAQGRAHWLKWQREQYEFDHALVYAEKNWGHSPFS PAAFPVFEPHWQICMAGGLSTGWIEWGGERFEFRDAPSYSEKNWGGGFPR EKFIKVT-----M-----KLPLTGQQYSEKVTENC
slr1737_SYNSP_S74814_ slr1737_ARATH_T04448_ CFI_ARATH_P41088_	RWFNLQANYFPDHPG-LSVTAAGGERIVLGRPE---EVALIGLHHQGNFY KWFVWQCNVFEGATGEVALTAGGGLRQLPGLTETVYENALVCVHYDGKMY VAIWKQLGLYTDCEA-KAV-----EKFLEIFKE---ET-----
slr1737_SYNSP_S74814_ slr1737_ARATH_T04448_ CFI_ARATH_P41088_	ETGPGHGTVTWQVAPWGRWQLKASNDRYWVKLSGKTDKKGSLVHTP-TAQ EFVFWNGVVRWEMSPWGYWYITAENENHVVELEARTNEAGTPLRAPTEV -FPPG-SSILFALSPTGSLTVAFSKDDS-IPETGIAVIENKLLAEA-VLE
slr1737_SYNSP_S74814_ slr1737_ARATH_T04448_ CFI_ARATH_P41088_	GLQLNCRDTRGYLYLQLGSVGHG-----LIVQGETDTAGLEVGG----- GLATACRDSCYGELKLQIWERLYDGSKGKVILETKSSMAAVEIGGGPWFG --SIIGKNGVSPGTRLSVAERLSQ-----LMMKNKDEKEVSDHSL-----
slr1737_SYNSP_S74814_ slr1737_ARATH_T04448_ CFI_ARATH_P41088_	----DWGLTEENLSKKT-----VPF----- TWKGDTSNTPELLKQALQVPLDLESALGLVPFFKPPGL ----EEKLAKEN-----

Figure 35

SEQUENCE LISTING

<110> Lassner, M
5 Post-Beittenmiller, D
Savidge, B
Weiss, J

<120> Nucleic Acid Sequences Involved in
10 Tocopherol Synthesis

<130> 17133/02/US

<150> 60/129,899
5 <151> 1999-04-15

<150> 60/146,461
<151> 1999-07-30

20 <160> 94

<170> FastSEQ for Windows Version 4.0

<210> 1
25 <211> 1182
<212> DNA
<213> Arabidopsis sp

<400> 1

30 atggagtcctc tgctctctag ttcttctctt gtttccgctg ctggtgggtt ttgttggaag 60
aagcagaatc taaagctcca ctctttatca gaaatccgag ttctgcgttg tgattcgagt 120
aaagttgtcg caaaaccgaa gtttaggaac aatcttgta ggctgatgg tcaaggatct 180
tcattgttgt tgtatccaaa acataagtcg agatttcggg ttaatgccac tgcgggctcag 240
cctgaggctt tcgactcgaa tagcaaacag aagtctttta gagactcgtt agatgcgttt 300
35 tacaggtttt ctaggcctca tacagttatt ggacacagtgc ttagcatttt atctgtatct 360
ttcttagcag tagagaaggt ttctgatata tctcctttac ttttcaactgg catcttggaag 420
gtgtgtgttg cagctctcat gatgaacatt tacatagttg ggctaaatca gttgtctgat 480
gttgaaatag ataaggttaa caagccctat cttccattgg catcaggaga atattctgtt 540
aacaccggca ttgcaatagt agcttccttc tccatcatga gtttctggct tgggtggatt 600
40 gttgggttcat ggccattgtt ctgggctctt tttgtgagtt tcatgctcgg tactgcatac 660

tctatcaatt tgccactttt acggtggaaa agatttgcac tggttgcagc aatgtgtatc 720
ctcgctgtcc gagctattat tgttcaaata gcctttttatc tacatattca gacacatgtg 780
tttggaagac caatcttggt cactaggcct cttatttttcg ccactgcgtt tatgagcttt 840
ttctctgtcg ttattgcatt gtttaaggat atacctgata tcgaagggga taagatatcc 900
5 ggaatccgat cattctctgt aactctgggt cagaaacggg tgttttggac atgtgtttaca 960
ctacttcaaa tggcttacgc tgttgcaatt ctagttggag ccacatctcc attcatatgg 1020
agcaaagtca tctcgggtgt gggtcattgt atactcgcaa caactttgtg ggctcgagct 1080
aagtcggtg atctgagtag caaaaccgaa ataacttcat gttatatgtt catatggaag 1140
ctcttttatg cagagtactt gctgttacct tttttgaagt ga 1182

<210> 2

<211> 393

<212> PRT

<213> Arabidopsis sp

<400> 2

Met Glu Ser Leu Leu Ser Ser Ser Ser Leu Val Ser Ala Ala Gly Gly
1 5 10 15
Phe Cys Trp Lys Lys Gln Asn Leu Lys Leu His Ser Leu Ser Glu Ile
20 25 30
Arg Val Leu Arg Cys Asp Ser Ser Lys Val Val Ala Lys Pro Lys Phe
35 40 45
Arg Asn Asn Leu Val Arg Pro Asp Gly Gln Gly Ser Ser Leu Leu Leu
50 55 60
Tyr Pro Lys His Lys Ser Arg Phe Arg Val Asn Ala Thr Ala Gly Gln
65 70 75 80
Pro Glu Ala Phe Asp Ser Asn Ser Lys Gln Lys Ser Phe Arg Asp Ser
85 90 95
Leu Asp Ala Phe Tyr Arg Phe Ser Arg Pro His Thr Val Ile Gly Thr
100 105 110
Val Leu Ser Ile Leu Ser Val Ser Phe Leu Ala Val Glu Lys Val Ser
115 120 125
Asp Ile Ser Pro Leu Leu Phe Thr Gly Ile Leu Glu Ala Val Val Ala
130 135 140
35 Ala Leu Met Met Asn Ile Tyr Ile Val Gly Leu Asn Gln Leu Ser Asp
145 150 155 160
Val Glu Ile Asp Lys Val Asn Lys Pro Tyr Leu Pro Leu Ala Ser Gly
165 170 175
Glu Tyr Ser Val Asn Thr Gly Ile Ala Ile Val Ala Ser Phe Ser Ile
40 180 185 190
Met Ser Phe Trp Leu Gly Trp Ile Val Gly Ser Trp Pro Leu Phe Trp

195 200 205
Ala Leu Phe Val Ser Phe Met Leu Gly Thr Ala Tyr Ser Ile Asn Leu
210 215 220
Pro Leu Leu Arg Trp Lys Arg Phe Ala Leu Val Ala Ala Met Cys Ile
5 225 230 235 240
Leu Ala Val Arg Ala Ile Ile Val Gln Ile Ala Phe Tyr Leu His Ile
245 250 255
Gln Thr His Val Phe Gly Arg Pro Ile Leu Phe Thr Arg Pro Leu Ile
260 265 270
10 Phe Ala Thr Ala Phe Met Ser Phe Phe Ser Val Val Ile Ala Leu Phe
275 280 285
Lys Asp Ile Pro Asp Ile Glu Gly Asp Lys Ile Phe Gly Ile Arg Ser
290 295 300
Phe Ser Val Thr Leu Gly Gln Lys Arg Val Phe Trp Thr Cys Val Thr
15 305 310 315 320
Leu Leu Gln Met Ala Tyr Ala Val Ala Ile Leu Val Gly Ala Thr Ser
325 330 335
Pro Phe Ile Trp Ser Lys Val Ile Ser Val Val Gly His Val Ile Leu
340 345 350
20 Ala Thr Thr Leu Trp Ala Arg Ala Lys Ser Val Asp Leu Ser Ser Lys
355 360 365
Thr Glu Ile Thr Ser Cys Tyr Met Phe Ile Trp Lys Leu Phe Tyr Ala
370 375 380
Glu Tyr Leu Leu Leu Pro Phe Leu Lys
25 385 390
<210> 3
<211> 1224
<212> DNA
30 <213> Arabidopsis sp

<400> 3
atggcggtttt ttggggtctc ccgtgtttca agacggttgt tgaaatcttc cgtctccgta 60
actccatctt ctccctctgc tcttttgcaa tcacaacata aatccttgtc caatcctgtg 120
35 actaccatt acacaaatcc tttcactaag tgttatcctt catggaatga taattaccaa 180
gtatggagta aaggaagaga attgcatcag gagaagtttt ttggtgttg ttggaattac 240
agattaattt gtggaatgtc gtcgtcttct tcggttttgg agggaaagcc gaagaaagat 300
gataaggaga agagtgatgg tgttggtgtt aagaaagctt cttggataga tttgtattta 360
ccagaagaag ttagaggtta tgctaagctt gctcgattgg ataaacccat tggaacttgg 420
40 ttgcttgctg gcccttgat gtggtcgatt gcgttggctg ctgatactgg aagccttcca 480
agttttaa atattggcttt atttggttgc ggagcattac ttcttagagg tgctggttgt 540

actataaatg atctgcttga tcaggacata gatacaaagg ttgatcgtac aaaactaaga 600
cctatcgcca gtggtctttt gacaccattt caagggattg gatttctcgg gctgcagttg 660
cttttaggct tagggattct tctccaactt aacaattaca gccgtgtttt aggggcttca 720
tctttgttac ttgtcttttc ctaccactt atgaagaggt ttacattttg gcctcaagcc 780
5 tttttagggt tgaccataaa ctggggagca ttgttaggat ggactgcagt taaaggaagc 840
atagcaccat ctattgtact ccctctctat ctctccggag tctgctggac ccttgtttat 900
gatactatct atgcacatca ggacaaagaa gatgatgtaa aagttggtgt taagtcaaca 960
gcccttagat tcggtgataa tacaaagctt tggttaactg gatttggcac agcatccata 1020
ggttttcttg cactttcttg attcagtgca gatctcgggt ggcaatatta cgcactcactg 1080
10 gccgctgcat caggacagtt aggatggcaa atagggacag ctgacttata atctggtgct 1140
gactgcagta gaaaatttgt gtcgaacaag tggtttggtg ctattatatt tagtggagtt 1200
gtacttgga gaagttttca ataa 1224

<210> 4

<211> 407

<212> PRT

<213> Arabidopsis sp

<400> 4

15 Met Ala Phe Phe Gly Leu Ser Arg Val Ser Arg Arg Leu Leu Lys Ser
1 5 10 15
Ser Val Ser Val Thr Pro Ser Ser Ser Ser Ala Leu Leu Gln Ser Gln
20 25 30
His Lys Ser Leu Ser Asn Pro Val Thr Thr His Tyr Thr Asn Pro Phe
25 35 40 45
Thr Lys Cys Tyr Pro Ser Trp Asn Asp Asn Tyr Gln Val Trp Ser Lys
50 55 60
Gly Arg Glu Leu His Gln Glu Lys Phe Phe Gly Val Gly Trp Asn Tyr
65 70 75 80
30 Arg Leu Ile Cys Gly Met Ser Ser Ser Ser Ser Val Leu Glu Gly Lys
85 90 95
Pro Lys Lys Asp Asp Lys Glu Lys Ser Asp Gly Val Val Val Lys Lys
100 105 110
Ala Ser Trp Ile Asp Leu Tyr Leu Pro Glu Glu Val Arg Gly Tyr Ala
35 115 120 125
Lys Leu Ala Arg Leu Asp Lys Pro Ile Gly Thr Trp Leu Leu Ala Trp
130 135 140
Pro Cys Met Trp Ser Ile Ala Leu Ala Ala Asp Pro Gly Ser Leu Pro
145 150 155 160
40 Ser Phe Lys Tyr Met Ala Leu Phe Gly Cys Gly Ala Leu Leu Leu Arg
165 170 175

Gly Ala Gly Cys Thr Ile Asn Asp Leu Leu Asp Gln Asp Ile Asp Thr
 180 185 190
 Lys Val Asp Arg Thr Lys Leu Arg Pro Ile Ala Ser Gly Leu Leu Thr
 195 200 205
 5 Pro Phe Gln Gly Ile Gly Phe Leu Gly Leu Gln Leu Leu Leu Gly Leu
 210 215 220
 Gly Ile Leu Leu Gln Leu Asn Asn Tyr Ser Arg Val Leu Gly Ala Ser
 225 230 235 240
 Ser Leu Leu Leu Val Phe Ser Tyr Pro Leu Met Lys Arg Phe Thr Phe
 10 245 250 255
 Trp Pro Gln Ala Phe Leu Gly Leu Thr Ile Asn Trp Gly Ala Leu Leu
 260 265 270
 Gly Trp Thr Ala Val Lys Gly Ser Ile Ala Pro Ser Ile Val Leu Pro
 275 280 285
 15 Leu Tyr Leu Ser Gly Val Cys Trp Thr Leu Val Tyr Asp Thr Ile Tyr
 290 295 300
 Ala His Gln Asp Lys Glu Asp Asp Val Lys Val Gly Val Lys Ser Thr
 305 310 315 320
 Ala Leu Arg Phe Gly Asp Asn Thr Lys Leu Trp Leu Thr Gly Phe Gly
 20 325 330 335
 Thr Ala Ser Ile Gly Phe Leu Ala Leu Ser Gly Phe Ser Ala Asp Leu
 340 345 350
 Gly Trp Gln Tyr Tyr Ala Ser Leu Ala Ala Ala Ser Gly Gln Leu Gly
 355 360 365
 25 Trp Gln Ile Gly Thr Ala Asp Leu Ser Ser Gly Ala Asp Cys Ser Arg
 370 375 380
 Lys Phe Val Ser Asn Lys Trp Phe Gly Ala Ile Ile Phe Ser Gly Val
 385 390 395 400
 Val Leu Gly Arg Ser Phe Gln
 30 405

<210> 5

<211> 1296

<212> DNA

35 <213> Arabidopsis sp

<400> 5

atgtggcgaa gatctgttgt ttctcgttta tcttcaagaa tctctgtttc ttcttcgtta 60
 ccaaacccta gactgattcc ttggtccgc gaattatgtg ccgttaatag cttctcccag 120
 40 cctccggtct cgacggaatc aactgctaag ttagggatca ctggtgtag atctgatgcc 180
 aatcgagttt ttgccactgc tactgccgcc gctacagcta cagctaccac cggtgagatt 240

tcgtctagag ttgcggcttt ggctggatta gggcatcact acgctcgttg ttattgggag 300
 ctttctaaag ctaaacttag tatgcttggt gttgcaactt ctggaactgg gtatattctg 360
 ggtacgggaa atgctgcaat tagcttcccg gggctttggt acacatgtgc aggaaccatg 420
 atgattgctg catctgctaa ttcttgaat cagatttttg agataagcaa tgattctaag 480
 5 atgaaaagaa cgatgctaag gccattgcct tcaggacgta ttagtggtcc acacgctggt 540
 gcatgggcta ctattgctgg tgcttctggt gcttggttggt tggccagcaa gactaatatg 600
 ttggctgctg gacttgcac tgccaatctt gtactttatg cgtttgttta tactccgttg 660
 aagcaacttc accctatcaa tacatgggtt ggcgctggtt ttggtgctat cccacccttg 720
 cttgggtggg cggcagcgtc tggtcagatt tcatacaatt cgatgattct tccagctgct 780
 10 ctttactttt ggcagatacc tcattttatg gcccttgcac atctctgccg caatgattat 840
 gcagctggag gttacaagat gttgtcactc tttgatccgt caggaagag aatagcagca 900
 gtggctctaa ggaactgctt ttacatgac cctctcggtt tcacgccta tgactggggg 960
 ttaacctcaa gttggttttg cctogaatca acacttctca cactagcaat cgctgcaaca 1020
 gcattttcat totaccgaga cgggaccatg cataaagcaa ggaaaatggt ccatgccagt 1080
 15 cttctcttcc ttctgtttt catgtctggt cttctctac accgtgtctc taatgataat 1140
 cagcaacaac tcgtagaaga agccggatta acaattctg tatctggtga agtcaaaact 1200
 cagaggcgaa agaaacgtgt ggcacaacct ccggtggctt atgcctctgc tgcaccgttt 1260
 cctttcctcc cagctccttc cttctactct ccatga 1296

 20 <210> 6
 <211> 431
 <212> PRT
 <213> Arabidopsis sp

 25 <400> 6
 Met Trp Arg Arg Ser Val Val Tyr Arg Phe Ser Ser Arg Ile Ser Val
 1 5 10 15
 Ser Ser Ser Leu Pro Asn Pro Arg Leu Ile Pro Trp Ser Arg Glu Leu
 20 25 30
 30 Cys Ala Val Asn Ser Phe Ser Gln Pro Pro Val Ser Thr Glu Ser Thr
 35 40 45
 Ala Lys Leu Gly Ile Thr Gly Val Arg Ser Asp Ala Asn Arg Val Phe
 50 55 60
 Ala Thr Ala Thr Ala Ala Thr Ala Thr Ala Thr Thr Gly Glu Ile
 35 65 70 75 80
 Ser Ser Arg Val Ala Ala Leu Ala Gly Leu Gly His His Tyr Ala Arg
 85 90 95
 Cys Tyr Trp Glu Leu Ser Lys Ala Lys Leu Ser Met Leu Val Val Ala
 100 105 110
 40 Thr Ser Gly Thr Gly Tyr Ile Leu Gly Thr Gly Asn Ala Ala Ile Ser
 115 120 125

Phe Pro Gly Leu Cys Tyr Thr Cys Ala Gly Thr Met Met Ile Ala Ala
 130 135 140
 Ser Ala Asn Ser Leu Asn Gln Ile Phe Glu Ile Ser Asn Asp Ser Lys
 145 150 155 160
 5 Met Lys Arg Thr Met Leu Arg Pro Leu Pro Ser Gly Arg Ile Ser Val
 165 170 175
 Pro His Ala Val Ala Trp Ala Thr Ile Ala Gly Ala Ser Gly Ala Cys
 180 185 190
 10 Leu Leu Ala Ser Lys Thr Asn Met Leu Ala Ala Gly Leu Ala Ser Ala
 195 200 205
 Asn Leu Val Leu Tyr Ala Phe Val Tyr Thr Pro Leu Lys Gln Leu His
 210 215 220
 Pro Ile Asn Thr Trp Val Gly Ala Val Val Gly Ala Ile Pro Pro Leu
 225 230 235 240
 15 Leu Gly Trp Ala Ala Ala Ser Gly Gln Ile Ser Tyr Asn Ser Met Ile
 245 250 255
 Leu Pro Ala Ala Leu Tyr Phe Trp Gln Ile Pro His Phe Met Ala Leu
 260 265 270
 Ala His Leu Cys Arg Asn Asp Tyr Ala Ala Gly Gly Tyr Lys Met Leu
 275 280 285
 20 Ser Leu Phe Asp Pro Ser Gly Lys Arg Ile Ala Ala Val Ala Leu Arg
 290 295 300
 Asn Cys Phe Tyr Met Ile Pro Leu Gly Phe Ile Ala Tyr Asp Trp Gly
 305 310 315 320
 25 Leu Thr Ser Ser Trp Phe Cys Leu Glu Ser Thr Leu Leu Thr Leu Ala
 325 330 335
 Ile Ala Ala Thr Ala Phe Ser Phe Tyr Arg Asp Arg Thr Met His Lys
 340 345 350
 Ala Arg Lys Met Phe His Ala Ser Leu Leu Phe Leu Pro Val Phe Met
 355 360 365
 30 Ser Gly Leu Leu Leu His Arg Val Ser Asn Asp Asn Gln Gln Gln Leu
 370 375 380
 Val Glu Glu Ala Gly Leu Thr Asn Ser Val Ser Gly Glu Val Lys Thr
 385 390 395 400
 35 Gln Arg Arg Lys Lys Arg Val Ala Gln Pro Pro Val Ala Tyr Ala Ser
 405 410 415
 Ala Ala Pro Phe Pro Phe Leu Pro Ala Pro Ser Phe Tyr Ser Pro
 420 425 430
 40 <210> 7
 <211> 479

<212> DNA

<213> Arabidopsis sp

<400> 7

5 ggaaactccc ggagcacctg tttgcaggta ccgctaacct taatcgataa tttattttctc 60
ttgtcaggaa ttatgtaagt ctggtggaag gctcgcatat cattttttgca ttgccttttcg 120
ctatgatcgg gtttacttttg ggtgtgatga gaccaggcgt ggcttttatgg tatggcgaaa 180
acccattttt atccaatgct gcattccctc ccgatgattc gttcttttcat tcctatacag 240
gtatcatgct gataaaaactg ttactgggtac tggtttgtat ggtatcagca agaagcgcg 300
10 cgatggcggt taaccggtat ctgcacaggc attttgacgc gaagaaccgc cgtactgccca 360
tccgtgaaat acctgcgggc gtcatatctg ccaacagtgc gctgggtgtt acgataggct 420
gctgcgtggt attctgggtg gcctgttatt tcattaacac gatctgtttt tacctggcg 479

<210> 8

<211> 551

<212> DNA

<213> Arabidopsis sp

<220>

<221> misc_feature

<222> (1)...(551)

<223> n = A,T,C or G

<400> 8

5 ttgtggctta caccttaatg agcatacgcc agnccattac ggctcgttaa tcggcgccat 60
ngccgngct gntgcaccgg tagtgaggta ctgcgccttg accaatcagc ttgatctagc 120
ggctcttatt ctgtttttta ttttactgtt ctggcaaagt ccgcattttt acgcgatttc 180
catttttcagg ctaaaagact tttcagcggc ctgtattccg gtgctgcca tcattaaaga 240
cctgcgctat accaaaatca gcatgctggt ttacgtgggc ttatttacac tggtctgctat 300
30 catgccggcc ctcttagggt atgccggttg gatztatggg atagcggcct taattttagg 360
cttgtatttg ctttatattg ccatacaagg attcaagacc gccgatgatc aaaaatggtc 420
tcgtaagatg tttggatctt cgattttaat cattaccctc ttgtcggtaa tgatgcttgt 480
ttaaacttac tgcctcctga agtttatata tcgataattt cagcttaagg aggcttagtg 540
gttaattcaa t 551

<210> 9

<211> 297

<212> PRT

<213> Arabidopsis sp

<400> 9

Met Val Leu Ala Glu Val Pro Lys Leu Ala Ser Ala Ala Glu Tyr Phe
1 5 10 15
Phe Lys Arg Gly Val Gln Gly Lys Gln Phe Arg Ser Thr Ile Leu Leu
20 25 30
5 Leu Met Ala Thr Ala Leu Asn Val Arg Val Pro Glu Ala Leu Ile Gly
35 40 45
Glu Ser Thr Asp Ile Val Thr Ser Glu Leu Arg Val Arg Gln Arg Gly
50 55 60
Ile Ala Glu Ile Thr Glu Met Ile His Val Ala Ser Leu Leu His Asp
10 65 70 75 80
Asp Val Leu Asp Asp Ala Asp Thr Arg Arg Gly Val Gly Ser Leu Asn
85 90 95
Val Val Met Gly Asn Lys Val Val Ala Leu Leu Ala Thr Ala Val Glu
100 105 110
45 His Leu Val Thr Gly Glu Thr Met Glu Ile Thr Ser Ser Thr Glu Gln
115 120 125
Arg Tyr Ser Met Asp Tyr Tyr Met Gln Lys Thr Tyr Tyr Lys Thr Ala
130 135 140
Ser Leu Ile Ser Asn Ser Cys Lys Ala Val Ala Val Leu Thr Gly Gln
20 145 150 155 160
Thr Ala Glu Val Ala Val Leu Ala Phe Glu Tyr Gly Arg Asn Leu Gly
165 170 175
Leu Ala Phe Gln Leu Ile Asp Asp Ile Leu Asp Phe Thr Gly Thr Ser
180 185 190
25 Ala Ser Leu Gly Lys Gly Ser Leu Ser Asp Ile Arg His Gly Val Ile
195 200 205
Thr Ala Pro Ile Leu Phe Ala Met Glu Glu Phe Pro Gln Leu Arg Glu
210 215 220
Val Val Asp Gln Val Glu Lys Asp Pro Arg Asn Val Asp Ile Ala Leu
30 225 230 235 240
Glu Tyr Leu Gly Lys Ser Lys Gly Ile Gln Arg Ala Arg Glu Leu Ala
245 250 255
Met Glu His Ala Asn Leu Ala Ala Ala Ala Ile Gly Ser Leu Pro Glu
260 265 270
35 Thr Asp Asn Glu Asp Val Lys Arg Ser Arg Arg Ala Leu Ile Asp Leu
275 280 285
Thr His Arg Val Ile Thr Arg Asn Lys
290 295
40 <210> 10
<211> 561

<212> DNA

<213> Arabidopsis sp

<400> 10

5	aagcgcaccc gtcctcttct acgattgccg ccagccgcat gtatggctgc ataaccgacc	60
	gccccatccc gctcgcggcc gcggtcgaat tcattcacac cgcgacgctg ctgcatgacg	120
	acgtcgtcga tgaaagcgat ttgcgcgcgc gccgcgaaag cgcgcataag gttttcggca	180
	atcaggcgag cgtgctcgtc ggcgatttcc ttttctcccg cgccttcag ctgatggtgg	240
	aagacggctc gctcgacgcg ctgcgcattc tctcggatgc ctccgccgtg atcgcgcagg	300
10	gcgaagtgat gcagctcggc accgcgcgca atcttgaaac caatatgagc cagtatctcg	360
	atgtgatcag cgcgaagacc gccgcgctct ttgcgcgcgc ctgcgaaatc ggcccgggtga	420
	tggcgaacgc gaaggcggaa gatgctgccg cgatgtgcga atacggcatg aatctcggta	480
	tcgccttcca gatcatcgac gaccttctcg attacggcac cggcggccac gccgagcttg	540
	gcaagaacac gggcgacgat t	561

<210> 11

<211> 966

<212> DNA

<213> Arabidopsis sp

<400> 11

15	atggtacttg ccgaggttcc aaagcttgcc tctgctgctg agtacttctt caaaaggggt	60
	gtgcaaggaa aacagtctcg ttcaactatt ttgctgctga tggcgacagc tctgaatgta	120
	cgcgttccag aagcattgat tggggaatca acagatatag tcacatcaga attacgcgta	180
5	aggcaacggg gtattgctga aatcaactgaa atgatacacg tcgcaagtct actgcacgat	240
	gatgtcttgg atgatgccga tacaaggcgt ggtgttggtt ccttaaagtgt tgtaatgggt	300
	aacaagatgt cggatttagc aggagacttc ttgctctccc gggcttgttg ggctctcgct	360
	gctttaaaga acacagaggt tgtagcatta cttgcaactg ctgtagaaca tcttggtacc	420
	ggtgaaacca tggaaataac tagttcaacc gagcagcgtt atagtatgga ctactacatg	480
30	cagaagacat attataagac agcatcgcta atctctaaca gctgcaaagc tgttgccgtt	540
	ctcaactggac aaacagcaga agttgcctg ttagcttttg agtatgggag gaatctgggt	600
	ttagcattcc aattaataga cgacattctt gatttcacgg gcacatctgc ctctctcgga	660
	aagggatcgt tgctagatat tcgcatgga gtcataacag cccaatcct ctttgccatg	720
	gaagagtctc ctcaactacg cgaagttggt gatcaagttg aaaaagatcc taggaatggt	780
35	gacattgctt tagagtatct tgggaagagc aagggaatac agagggcaag agaattagcc	840
	atggaacatg cgaatctagc agcagctgca atcgggtctc tacctgaaac agacaatgaa	900
	gatgtcaaaa gatcgaggcg ggcacttatt gacttgacct atagagtcac caccagaaac	960
	aagtga	966

40 <210> 12

<211> 321

<212> PRT

<213> Arabidopsis sp

<400> 12

5 Met Val Leu Ala Glu Val Pro Lys Leu Ala Ser Ala Ala Glu Tyr Phe
1 5 10 15
Phe Lys Arg Gly Val Gln Gly Lys Gln Phe Arg Ser Thr Ile Leu Leu
20 25 30
Leu Met Ala Thr Ala Leu Asn Val Arg Val Pro Glu Ala Leu Ile Gly
10 35 40 45
Glu Ser Thr Asp Ile Val Thr Ser Glu Leu Arg Val Arg Gln Arg Gly
50 55 60
Ile Ala Glu Ile Thr Glu Met Ile His Val Ala Ser Leu Leu His Asp
65 70 75 80
5 Asp Val Leu Asp Asp Ala Asp Thr Arg Arg Gly Val Gly Ser Leu Asn
85 90 95
Val Val Met Gly Asn Lys Met Ser Val Leu Ala Gly Asp Phe Leu Leu
100 105 110
Ser Arg Ala Cys Gly Ala Leu Ala Ala Leu Lys Asn Thr Glu Val Val
115 120 125
Ala Leu Leu Ala Thr Ala Val Glu His Leu Val Thr Gly Glu Thr Met
130 135 140
Glu Ile Thr Ser Ser Thr Glu Gln Arg Tyr Ser Met Asp Tyr Tyr Met
145 150 155 160
25 Gln Lys Thr Tyr Tyr Lys Thr Ala Ser Leu Ile Ser Asn Ser Cys Lys
165 170 175
Ala Val Ala Val Leu Thr Gly Gln Thr Ala Glu Val Ala Val Leu Ala
180 185 190
Phe Glu Tyr Gly Arg Asn Leu Gly Leu Ala Phe Gln Leu Ile Asp Asp
30 195 200 205
Ile Leu Asp Phe Thr Gly Thr Ser Ala Ser Leu Gly Lys Gly Ser Leu
210 215 220
Ser Asp Ile Arg His Gly Val Ile Thr Ala Pro Ile Leu Phe Ala Met
225 230 235 240
35 Glu Glu Phe Pro Gln Leu Arg Glu Val Val Asp Gln Val Glu Lys Asp
245 250 255
Pro Arg Asn Val Asp Ile Ala Leu Glu Tyr Leu Gly Lys Ser Lys Gly
260 265 270
Ile Gln Arg Ala Arg Glu Leu Ala Met Glu His Ala Asn Leu Ala Ala
40 275 280 285
Ala Ala Ile Gly Ser Leu Pro Glu Thr Asp Asn Glu Asp Val Lys Arg

290 295 300
Ser Arg Arg Ala Leu Ile Asp Leu Thr His Arg Val Ile Thr Arg Asn
305 310 315 320
Lys

<210> 13

<211> 621

<212> DNA

<213> Arabidopsis sp

<400> 13

gctttctcct	ttgctaattc	ttgagctttc	ttgatccac	cgcgatttct	aactatttca	60
atcgcttctt	caagcgatcc	aggctcacia	aactcagact	caatgatctc	tcttagcctt	120
ggctcattct	ctagcgcgaa	gatcaactgg	gcggttatgt	tacctttggc	taagtcatta	180
gctgcaggct	tacctaactg	ctctgtggac	tgagtgaagt	ccagaatgtc	atcaactact	240
tgaaaagata	aaccgagatt	cttcccgaa	tgatacattt	gctctgcgac	cttgctttcg	300
actttactga	aaattgctgc	tcctttggtg	cttgacagct	ctaataaagt	tgtctttag	360
taactcttta	gcatgtagtc	atcaagcttg	acatcacaat	cgaataaact	cgatgcttgc	420
tttatctcac	cgcttgcaaa	atctttgatc	acctgcaaaa	agataaatca	agattcagac	480
caaagtctct	ttgtattgag	tagcttcac	taatctcaga	aaggaatatt	acctgactta	540
tgagcttaat	gacttcaagg	ttttcgagat	ttgtaagtac	catgatgctt	gagcaacatg	600
aaatccccag	ctaatacagc	t				621

<210> 14

<211> 741

<212> DNA

<213> Arabidopsis sp

<400> 14

ggtgagtttt	gttaatagtt	atgagattca	tctatttttg	tcataaaatt	gtttggtttg	60
gtttaaactc	tgtgtataat	tgaggaaag	gaaacagttc	atgagctttt	cggcacaaga	120
gtagcgggtg	tagctggaga	tttcatgttt	gctcaagcgt	catggtactt	agcaaacttc	180
gagaatcttg	aagttattaa	gtcatcagt	caggtactta	gttactctta	cattgttttt	240
ctatgagggt	gagctatgaa	tctcatttcg	ttgaataatg	ctgtgcctca	aacttttttt	300
catgttttca	ggtgatcaaa	gactttgcaa	gcggagagat	aaagcaggcg	tcagccttat	360
ttgactgcca	caccaagctc	gacgagtact	tactcaaaag	tttctacaag	acagcctctt	420
tagtggctgc	gagcaccaaa	ggagctgcca	ttttcagcag	agttgagcct	gatgtgacag	480
aacaaatgta	cgagtttggg	aagaatctcg	gtctctcttt	ccagatagtt	gatgatattt	540
tggatttcac	tcagtcgaca	gagcagctcg	ggaagccagc	agggagtgat	ttggctaaaag	600
gtaacttaac	agcacctgtg	attttcgctc	tgagagagga	gccaaggcta	agagagatca	660

ttgagtcaaa gttctgtgag gcggtttctc tggaagaagc gattgaagcg gtgacaaaag 720
gtgggggggat taagagagca c 741

<210> 15

5 <211> 1087

<212> DNA

<213> Arabidopsis sp

<400> 15

10 cctcttcagc caatccagag gaagaagaga caacttttta tctttcgtca agagtctccg 60
aaaacgcacg gttttatgct ctctcttctg ccctcacctc acaagacgca gggcacatga 120
ttcaaccaga gggaaaaagc aacgataaca actctgcttt tgatttcaag ctgtatatga 180
tccgcaaagc cgagtctgta aatgcggctc tcgacgtttc cgtaccgctt ctgaaacccc 240
ttacgatcca agaagcggtc aggtactctt tgctagccgg cggaaaaacgt gtgaggcctc 300
5 tgctctgcat tgccgcttgt gagcttgtgg ggggcgacga ggetactgcc atgtcagccg 360
cttgcgcggt cgagatgac cacaacagct ctctcattca tgacgatctt ccgtgcatgg 420
acaatgccga cctccgtaga ggcaagccca ccaatcaca ggatgttgt ttaattatat 480
gaaggctcag agataatgct gaactagtgt tgaaccaatt tttgctcaa caaggtatat 540
ggagaagaca tggcggtttt ggcagggtgat gcaactcctg cattggcggt tgagcacatg 600
0 acggttgtgt cgagtgggtt ggctgcctcc gagaagatga ttcgcgccgt ggttgagctg 660
gccagggccca tagggactac agggctagtt gctggacaaa tgatagacct agccagcgaa 720
agactgaatc cagacaaggt tggattggag catctagagt tcatccatct ccacaaaacg 780
gcggcattgt tggaggcagc ggcagtttta ggggttataa tgggaggtgg aacagaggaa 840
gaaatcgaaa agcttagaaa gtatgctagg tgtattggac tactgtttca ggttgttgat 900
5 gacattctcg acgtaacaaa atctactgag gaattgggtg agacagccgg aaaagacgta 960
atggccggaa agctgacgta tccaaggctg atagggttgg agggatccag ggaagttgca 1020
gagcacctga ggagagaagc agaggaaaag cttaaagggt ttgatccaag tcaggcggcg 1080
cctctgg 1087

30 <210> 16

<211> 1164

<212> DNA

<213> Arabidopsis sp

35 <400> 16

atgacttcga ttctcaacac tgtctccacc atccactctt ccagagttac ctccgtcgat 60
cgagtccgag tctctctct tgggaattcg gattccgttg agttcactcg ccggcggtct 120
ggtttctcga cgttgatcta cgaatcacc gggcgagat ttgttggtcg tgcggcggag 180
actgatactg ataaagttaa atctcagaca cctgacaagg caccagccgg tggttcaagc 240
40 attaaccagc ttctcgggtat caaaggagca tctcaagaaa ctaataaatg gaagattcgt 300
cttcagctta caaaaccagt cacttggcct ccaactgggtt ggggagtcgt ctgtggtgct 360

5 gctgcttcag ggaactttca ttggacccca gaggatgttg ctaagtcgat tctttgcatg 420
atgatgtctg gtccttgtct tactggctat acacagacaa tcaacgactg gtatgataga 480
gatatcgacg caattaatga gccatatcgt ccaattccat ctggagcaat atcagagcca 540
gaggttatta cacaagtctg ggtgctatta ttgggaggtc ttggtattgc tggaatatta 600
gatgtgtggg cagggcatac cactccact gtcttctatc ttgctttggg aggatcattg 660
ctatcttata tatactctgc tccacctctt aagctaaaac aaaatggatg ggttggaat 720
tttgacttg gagcaagcta tattagtttg ccatgggtggg ctggccaagc attgtttggc 780
actcttacgc cagatgttgt tgttctaaca ctcttgtaaca gcatagctgg gttaggaata 840
gccattgta acgaacttcaa aagtgttgaa ggagatagag cattaggact tcagtctctc 900
10 ccagtagctt ttggcaccga aactgcaaaa tggatatgcg ttggtgctat agacattact 960
cagctttctg ttgccggata tctattagca tctgggaaac cttattatgc gttggcggtg 1020
gttgctttga tcattcctca gattgtgttc cagtttaaat actttctcaa ggacctgtc 1080
aaatacgacg tcaagtacca ggcaagcgcg cagccattct tgggtgctcg aatatttgta 1140
acggcattag catcgcaaca ctga 1164

<210> 17

<211> 387

<212> PRT

<213> Arabidopsis sp

<400> 17

Met Thr Ser Ile Leu Asn Thr Val Ser Thr Ile His Ser Ser Arg Val

1 5 10 15

Thr Ser Val Asp Arg Val Gly Val Leu Ser Leu Arg Asn Ser Asp Ser

20 25 30

Val Glu Phe Thr Arg Arg Arg Ser Gly Phe Ser Thr Leu Ile Tyr Glu

35 40 45

Ser Pro Gly Arg Arg Phe Val Val Arg Ala Ala Glu Thr Asp Thr Asp

50 55 60

30 Lys Val Lys Ser Gln Thr Pro Asp Lys Ala Pro Ala Gly Gly Ser Ser

65 70 75 80

Ile Asn Gln Leu Leu Gly Ile Lys Gly Ala Ser Gln Glu Thr Asn Lys

85 90 95

Trp Lys Ile Arg Leu Gln Leu Thr Lys Pro Val Thr Trp Pro Pro Leu

100 105 110

35 Val Trp Gly Val Val Cys Gly Ala Ala Ala Ser Gly Asn Phe His Trp

115 120 125

Thr Pro Glu Asp Val Ala Lys Ser Ile Leu Cys Met Met Met Ser Gly

130 135 140

40 Pro Cys Leu Thr Gly Tyr Thr Gln Thr Ile Asn Asp Trp Tyr Asp Arg

145 150 155 160

Asp Ile Asp Ala Ile Asn Glu Pro Tyr Arg Pro Ile Pro Ser Gly Ala
165 170 175
Ile Ser Glu Pro Glu Val Ile Thr Gln Val Trp Val Leu Leu Leu Gly
180 185 190
5 Gly Leu Gly Ile Ala Gly Ile Leu Asp Val Trp Ala Gly His Thr Thr
195 200 205
Pro Thr Val Phe Tyr Leu Ala Leu Gly Gly Ser Leu Leu Ser Tyr Ile
210 215 220
Tyr Ser Ala Pro Pro Leu Lys Leu Lys Gln Asn Gly Trp Val Gly Asn
10 225 230 235 240
Phe Ala Leu Gly Ala Ser Tyr Ile Ser Leu Pro Trp Trp Ala Gly Gln
245 250 255
Ala Leu Phe Gly Thr Leu Thr Pro Asp Val Val Val Leu Thr Leu Leu
260 265 270
5 Tyr Ser Ile Ala Gly Leu Gly Ile Ala Ile Val Asn Asp Phe Lys Ser
275 280 285
Val Glu Gly Asp Arg Ala Leu Gly Leu Gln Ser Leu Pro Val Ala Phe
290 295 300
Gly Thr Glu Thr Ala Lys Trp Ile Cys Val Gly Ala Ile Asp Ile Thr
0 305 310 315 320
Gln Leu Ser Val Ala Gly Tyr Leu Leu Ala Ser Gly Lys Pro Tyr Tyr
325 330 335
Ala Leu Ala Leu Val Ala Leu Ile Ile Pro Gln Ile Val Phe Gln Phe
340 345 350
5 Lys Tyr Phe Leu Lys Asp Pro Val Lys Tyr Asp Val Lys Tyr Gln Ala
355 360 365
Ser Ala Gln Pro Phe Leu Val Leu Gly Ile Phe Val Thr Ala Leu Ala
370 375 380
Ser Gln His
30 385

<210> 18

<211> 981

<212> DNA

35 <213> Arabidopsis sp

<400> 18

atgttggttta gtggttcagc gatcccattha agcagcttct gctctcttcc ggagaaaccc 60
cacactcttc ctatgaaact ctctcccgct gcaatccgat cttcatcctc atctgccccg 120
40 gggtcggttg acttcgatct gaggaagtat tggacgactc tgatcaccga gatcaaccag 180
aagctggatg aggccataacc ggtcaagcac cctgcgggga tctacgaggc tatgagatac 240

	tctgtactcg cacaaggcgc caagcgtgcc cctcctgtga tgtgtgtggc ggccctgcgag	300
	ctcttcgggtg gcgatcgccct cgccgctttc cccaccgcct gtgccctaga aatggtgcac	360
	gcggtcttcgt tgatacacga cgacctcccc tgtatggacg acgatcctgt gcgcagagga	420
	aagccatcta accacactgt ctacggctct ggcatggcca ttctcgccgg tgacgcacctc	480
5	ttcccactcg ccttccagca cattgtctcc cacacgcctc ctgaccttgt tccccgagcc	540
	accatcctca gactcatcac tgagattgcc cgcaactgtcg gctccactgg tatggctgca	600
	ggccagtagc tcgaccttga aggaggtccc tttcctcttt cctttgttca ggagaagaaa	660
	ttcggagcca tgggtgaatg ctctgccgtg tgcggtggcc tattgggcgg tgccactgag	720
	gatgagctcc agagtctccg aaggtacggg agagccgtcg ggatgctgta tcaggtggtc	780
10	gatgacatca ccgaggacaa gaagaagagc tatgatgggtg gagcagagaa gggaatgatg	840
	gaaatggcgg aagagctcaa ggagaaggcg aagaaggagc ttcaagtgtt tgacaacaag	900
	tatggaggag gagacacact tgttcctctc tacaccttcg ttgactacgc tgctcatoga	960
	cattttcttc ttcccctctg a	981

5 <210> 19
 <211> 245
 <212> DNA
 <213> Glycine sp

20 <400> 19
 gcaacatctg ggactgggtt tgtcttgggg agtggtagtg ctgttgatct ttccggcattt 60
 tcttgcaatt gcttgggtac catgatggtt gctgcatctg ctaactcttt gaatcaggtg 120
 tttgagatca ataattgatgc taaaatgaag agaacaagtc gcaggccact acctcagga 180
 cgcatacaca tacctcatgc agttggctgg gcatactctg ttggattagc tggtagcgct 240
 ctact 245

30 <210> 20
 <211> 253
 <212> DNA
 <213> Glycine sp

35 <400> 20
 attggctttc caagatcatt ggggttttctt gttgcattca tgaccttcta ctcttgggt 60
 ttggcattgt ccaaggatat acctgacgtt gaaggagata aagagcacgg cattgattct 120
 tttgcagtag gtctaggtca gaaacgggca ttttgatttt gcgtttcctt ttttgaaatg 180
 gctttcggag ttggtatcct ggccggagca tcatgctcac acttttggac taaaattttc 240
 acgggtatgg gaa 253

40 <210> 21
 <211> 275
 <212> DNA

<213> Glycine sp

<400> 21

5 tgatcttcta ctctctgggt atggcattgt ccaaggatat atctgacgtt aaaggagata 60
aagcatacgg catcgatact ttagcgatac gtttgggtca aaaatgggta ttttggattt 120
gcattatcct ttttgaaatg gcttttggag ttgccctctt ggcaggagca acatcttctt 180
acctttggat taaaattgtc acgggtctgg gacatgctat tcttgcttca attctcttgt 240
accaagccaa atctatatac ttgagcaaca aagtt 275

10 <210> 22

<211> 299

<212> DNA

<213> Glycine sp

15 <220>

<221> misc_feature

<222> (1)...(299)

<223> n = A,T,C or G

20 <400> 22

ccanaatang tncatcttng aaagacaatt ggctctttca acacacaagt ctgcatgtga 60
agaagaggcc aattgtcttt ccaagatcac ttatngtggc tattgtaatc atgaacttct 120
tctttgtggg tatggcattg gcaaaggata tacctanctg ttgaaggaga taaaatatat 180
ggcattgata cttttgcaat acgtataggt caaaaacaag ttttttggat ttgtattttc 240
ctttttgaaa ggctttcgga gtttccctag tggcaggagc aacatcttct agccttggt 299

<210> 23

<211> 767

<212> DNA

30 <213> Glycine sp

<400> 23

gtggaggctg tggttgctgc cctgtttatg aatattttata ttgttggttt gaatcaattg 60
tctgatgttg aaatagacaa gataaacaag ccgtatcttc cattagcatc tggggaatat 120
35 tcctttgaaa ctgggtgtcac tattgttgca tctttttcaa ttctgagttt ttggcttggc 180
tgggttgtag gttcatggcc attatttttg gccctttttg taagctttgt gctaggaact 240
gcttattcaa tcaatgtgcc tctgttgaga tggaagaggt ttgcagtgtc tgcagcgatg 300
tgcattctag ctgttcgggc agtaatagtt caacttgcac ttttccttca catgcagact 360
catgtgtaca agaggccaac tgtottttca agaccattga tttttgctac tgcattcatg 420
40 agcttcttct ctgtagttat agcactgttt aaggatatac ctgacattga aggagataaa 480
gtattttggca tccaatcttt ttcagtgtgt ttaggtcaga agccggtgtt ctggacttgt 540

gttacccttc ttgaaatagc ttatggagtc gccctcctgg tgggagctgc atctccttgt 600
ctttggagca aaattttcac gggctctggga cacgctgtgc tggcttcaat tctctggttt 660
catgccaaat ctgtagattt gaaaagcaaa gottcgataa catccttcta tatgtttatt 720
tggaagctat tttatgcaga atacttactc attccttttg ttagatg 767

5

<210> 24
<211> 255
<212> PRT
<213> Glycine sp

10

<400> 24

Val Glu Ala Val Val Ala Ala Leu Phe Met Asn Ile Tyr Ile Val Gly
1 5 10 15
Leu Asn Gln Leu Ser Asp Val Glu Ile Asp Lys Ile Asn Lys Pro Tyr
20 25 30
Leu Pro Leu Ala Ser Gly Glu Tyr Ser Phe Glu Thr Gly Val Thr Ile
35 40 45
Val Ala Ser Phe Ser Ile Leu Ser Phe Trp Leu Gly Trp Val Val Gly
50 55 60
Ser Trp Pro Leu Phe Trp Ala Leu Phe Val Ser Phe Val Leu Gly Thr
65 70 75 80
Ala Tyr Ser Ile Asn Val Pro Leu Leu Arg Trp Lys Arg Phe Ala Val
85 90 95
Leu Ala Ala Met Cys Ile Leu Ala Val Arg Ala Val Ile Val Gln Leu
100 105 110
Ala Phe Phe Leu His Met Gln Thr His Val Tyr Lys Arg Pro Pro Val
115 120 125
Phe Ser Arg Pro Leu Ile Phe Ala Thr Ala Phe Met Ser Phe Phe Ser
130 135 140
Val Val Ile Ala Leu Phe Lys Asp Ile Pro Asp Ile Glu Gly Asp Lys
145 150 155 160
Val Phe Gly Ile Gln Ser Phe Ser Val Cys Leu Gly Gln Lys Pro Val
165 170 175
Phe Trp Thr Cys Val Thr Leu Leu Glu Ile Ala Tyr Gly Val Ala Leu
180 185 190
Leu Val Gly Ala Ala Ser Pro Cys Leu Trp Ser Lys Ile Phe Thr Gly
195 200 205
Leu Gly His Ala Val Leu Ala Ser Ile Leu Trp Phe His Ala Lys Ser
210 215 220
Val Asp Leu Lys Ser Lys Ala Ser Ile Thr Ser Phe Tyr Met Phe Ile
225 230 235 240

Trp Lys Leu Phe Tyr Ala Glu Tyr Leu Leu Ile Pro Phe Val Arg
245 250 255

<210> 25

5 <211> 360

<212> DNA

<213> Zea sp

<220>

10 <221> misc_feature

<222> (1)...(360)

<223> n = A,T,C or G

<400> 25

5 ggcgtcttca cttgttctgg tcttctcgta tcccctgatg aagaggttca cattttggcc 60
tcaggcttat cttggcctga cattcaactg gggagcttta ctagggtggg ctgctattaa 120
ggaaagcata gaccctgcaa atcatccttc cattgtatac agctgggtatt tgttggacgc 180
tggtgtatga tactatatat gcgcacacagg tgtttcgcata tccctacttt catattaatc 240
cttgatgaag tggccatttc atgttgctgc ggtggtctta tacttgcata tctccatgca 300
0 tctcaggaca aagangatga cctgaaagta ggagtccaag tccacagctt aagatttggg 360

<210> 26

<211> 299

<212> DNA

5 <213> Zea sp

<220>

<221> misc_feature

<222> (1)...(299)

30 <223> n = A,T,C or G

<400> 26

gatggttgca gcatctgcaa ataccctcaa ccagggtgttt gngataaaaa atgatgctaa 60
aatgaaaagg acaatgcgtg ccccctgcca tctggtcgca ttagtctctgc acatgctgcg 120
35 atgtgggcta caagtgttgg agttgcagga acagctttgt tggcctggaa ggctaattgc 180
ttggcagctg ggcttgacgc ttctaattctt gttctgtatg catttggtgta tacgccgttg 240
aagcaaatac accctgttaa tacatggggtt ggggcagtcg ttggtgccat cccaccact 299

<210> 27

40 <211> 255

<212> DNA

<213> Zea sp

<220>

<221> misc_feature

5 <222> (1)...(255)

<223> n = A,T,C or G

<400> 27

10 anacttgcat atctccatgc ntctcaggac aaagangatg acctgaaagt aggtgtcaag 60
tccacagcat taagatttgg agatttgacc nnatactgna tcagtggctt tggcgcggca 120
tgcttcggca gcttagcact cagtggttac aatgctgacc ttggttggtg tttagtgtga 180
tgcttgagcg aagaatggta tngtttttac ttgatattga ctccagacct gaaatcatgt 240
tggaacagggt ggccc 255

5 <210> 28

<211> 257

<212> DNA

<213> Zea sp

0 <400> 28

attgaagggg ataggactct ggggcttcag tcaacttctg ttgcttttgg gatggaaact 60
gcaaaatgga tttgtgttgg agcaattgat atcactcaat tatctgttgc aggttaccta 120
ttgagcaccg gtaagctgta ttatgccctg gtgttgcttg ggctaacaat tctcagggtg 180
ttctttcagt tccagtactt cctgaaggac cctgtgaagt atgatgtcaa atatcaggca 240
5 agcgcacaaac cattctt 257

<210> 29

<211> 368

<212> DNA

30 <213> Zea sp

<400> 29

atccagttgc aaataataat ggcgttcttc tctgttgtaa tagcactatt caaggatata 60
cctgacatcg aaggggaccg catattcggg atccgatcct tcagegtccg gttagggcaa 120
35 aagaaggtct tttggatctg cgttggtctg cttgagatgg cctacagcgt tgcgatactg 180
atgggagcta cctcttcctg tttgtggagc aaaacagcaa ccatcgctgg ccattccata 240
cttgccgcga tcctatggag ctgcgcgcga tcggtggact tgacgagcaa agccgcaata 300
acgtccttct acatgttcat ctggaagctg ttctacgcgg agtacctgct catccctctg 360
gtgcggtg 368

40 <210> 30

<211> 122
<212> PRT
<213> Zea sp

5 <400> 30

Ile Gln Leu Gln Ile Ile Met Ala Phe Phe Ser Val Val Ile Ala Leu
1 5 10 15
Phe Lys Asp Ile Pro Asp Ile Glu Gly Asp Arg Ile Phe Gly Ile Arg
20 25 30
10 Ser Phe Ser Val Arg Leu Gly Gln Lys Lys Val Phe Trp Ile Cys Val
35 40 45
Gly Leu Leu Glu Met Ala Tyr Ser Val Ala Ile Leu Met Gly Ala Thr
50 55 60
Ser Ser Cys Leu Trp Ser Lys Thr Ala Thr Ile Ala Gly His Ser Ile
65 70 75 80
Leu Ala Ala Ile Leu Trp Ser Cys Ala Arg Ser Val Asp Leu Thr Ser
85 90 95
Lys Ala Ala Ile Thr Ser Phe Tyr Met Phe Ile Trp Lys Leu Phe Tyr
100 105 110
20 Ala Glu Tyr Leu Leu Ile Pro Leu Val Arg
115 120

<210> 31
<211> 278
<212> DNA
<213> Zea sp

<400> 31

tattcagcac cacctotcaa gctcaagcag aatggatgga ttgggaactt cgctctgggt 60
30 gcgagttaca tcagcttgcc ctggtgggct ggccaggcgt tatttggaac tcttacacca 120
gatatcattg tottgactac tttgtacagc atagctgggc tagggattgc tattgtaa 180
gatttcaaga gtattgaagg ggaataggact ctggggcttc agtcacttcc tgttgctttt 240
gggatggaaa ctgcaaaatg gatttgtgtt ggagcaat 278

35 <210> 32
<211> 292
<212> PRT
<213> Synechocystis sp

40 <400> 32

Met Val Ala Gln Thr Pro Ser Ser Pro Pro Leu Trp Leu Thr Ile Ile

1 5 10 15
Tyr Leu Leu Arg Trp His Lys Pro Ala Gly Arg Leu Ile Leu Met Ile
20 25 30
Pro Ala Leu Trp Ala Val Cys Leu Ala Ala Gln Gly Leu Pro Pro Leu
5 35 40 45
Pro Leu Leu Gly Thr Ile Ala Leu Gly Thr Leu Ala Thr Ser Gly Leu
50 55 60
Gly Cys Val Val Asn Asp Leu Trp Asp Arg Asp Ile Asp Pro Gln Val
65 70 75 80
10 Glu Arg Thr Lys Gln Arg Pro Leu Ala Ala Arg Ala Leu Ser Val Gln
85 90 95
Val Gly Ile Gly Val Ala Leu Val Ala Leu Leu Cys Ala Ala Gly Leu
100 105 110
Ala Phe Tyr Leu Thr Pro Leu Ser Phe Trp Leu Cys Val Ala Ala Val
115 120 125
Pro Val Ile Val Ala Tyr Pro Gly Ala Lys Arg Val Phe Pro Val Pro
130 135 140
Gln Leu Val Leu Ser Ile Ala Trp Gly Phe Ala Val Leu Ile Ser Trp
145 150 155 160
20 Ser Ala Val Thr Gly Asp Leu Thr Asp Ala Thr Trp Val Leu Trp Gly
165 170 175
Ala Thr Val Phe Trp Thr Leu Gly Phe Asp Thr Val Tyr Ala Met Ala
180 185 190
Asp Arg Glu Asp Asp Arg Arg Ile Gly Val Asn Ser Ser Ala Leu Phe
195 200 205
Phe Gly Gln Tyr Val Gly Glu Ala Val Gly Ile Phe Phe Ala Leu Thr
210 215 220
Ile Gly Cys Leu Phe Tyr Leu Gly Met Ile Leu Met Leu Asn Pro Leu
225 230 235 240
30 Tyr Trp Leu Ser Leu Ala Ile Ala Ile Val Gly Trp Val Ile Gln Tyr
245 250 255
Ile Gln Leu Ser Ala Pro Thr Pro Glu Pro Lys Leu Tyr Gly Gln Ile
260 265 270
Phe Gly Gln Asn Val Ile Ile Gly Phe Val Leu Leu Ala Gly Met Leu
35 275 280 285
Leu Gly Trp Leu
290

<210> 33

40 <211> 316

<212> PRT

<213> Synechocystis sp

<400> 33

5 Met Val Thr Ser Thr Lys Ile His Arg Gln His Asp Ser Met Gly Ala
1 5 10 15
Val Cys Lys Ser Tyr Tyr Gln Leu Thr Lys Pro Arg Ile Ile Pro Leu
20 25 30
Leu Leu Ile Thr Thr Ala Ala Ser Met Trp Ile Ala Ser Glu Gly Arg
35 40 45
10 Val Asp Leu Pro Lys Leu Leu Ile Thr Leu Leu Gly Gly Thr Leu Ala
50 55 60
Ala Ala Ser Ala Gln Thr Leu Asn Cys Ile Tyr Asp Gln Asp Ile Asp
65 70 75 80
Tyr Glu Met Leu Arg Thr Arg Ala Arg Pro Ile Pro Ala Gly Lys Val
5 85 90 95
Gln Pro Arg His Ala Leu Ile Phe Ala Leu Ala Leu Gly Val Leu Ser
100 105 110
Phe Ala Leu Leu Ala Thr Phe Val Asn Val Leu Ser Gly Cys Leu Ala
115 120 125
20 Leu Ser Gly Ile Val Phe Tyr Met Leu Val Tyr Thr His Trp Leu Lys
130 135 140
Arg His Thr Ala Gln Asn Ile Val Ile Gly Gly Ala Ala Gly Ser Ile
145 150 155 160
Pro Pro Leu Val Gly Trp Ala Ala Val Thr Gly Asp Leu Ser Trp Thr
5 165 170 175
Pro Trp Val Leu Phe Ala Leu Ile Phe Leu Trp Thr Pro Pro His Phe
180 185 190
Trp Ala Leu Ala Leu Met Ile Lys Asp Asp Tyr Ala Gln Val Asn Val
195 200 205
30 Pro Met Leu Pro Val Ile Ala Gly Glu Glu Lys Thr Val Ser Gln Ile
210 215 220
Trp Tyr Tyr Ser Leu Leu Val Val Pro Phe Ser Leu Leu Leu Val Tyr
225 230 235 240
Pro Leu His Gln Leu Gly Ile Leu Tyr Leu Ala Ile Ala Ile Ile Leu
35 245 250 255
Gly Gly Gln Phe Leu Val Lys Ala Trp Gln Leu Lys Gln Ala Pro Gly
260 265 270
Asp Arg Asp Leu Ala Arg Gly Leu Phe Lys Phe Ser Ile Phe Tyr Leu
275 280 285
40 Met Leu Leu Cys Leu Ala Met Val Ile Asp Ser Leu Pro Val Thr His
290 295 300

Gln Leu Val Ala Gln Met Gly Thr Leu Leu Leu Gly
305 310 315

<210> 34

5 <211> 324

<212> PRT

<213> Synechocystis sp

<400> 34

10 Met Ser Asp Thr Gln Asn Thr Gly Gln Asn Gln Ala Lys Ala Arg Gln
1 5 10 15
Leu Leu Gly Met Lys Gly Ala Ala Pro Gly Glu Ser Ser Ile Trp Lys
20 25 30
Ile Arg Leu Gln Leu Met Lys Pro Ile Thr Trp Ile Pro Leu Ile Trp
35 40 45
Gly Val Val Cys Gly Ala Ala Ser Ser Gly Gly Tyr Ile Trp Ser Val
50 55 60
Glu Asp Phe Leu Lys Ala Leu Thr Cys Met Leu Leu Ser Gly Pro Leu
65 70 75 80
20 Met Thr Gly Tyr Thr Gln Thr Leu Asn Asp Phe Tyr Asp Arg Asp Ile
85 90 95
Asp Ala Ile Asn Glu Pro Tyr Arg Pro Ile Pro Ser Gly Ala Ile Ser
100 105 110
Val Pro Gln Val Val Thr Gln Ile Leu Ile Leu Leu Val Ala Gly Ile
115 120 125
Gly Val Ala Tyr Gly Leu Asp Val Trp Ala Gln His Asp Phe Pro Ile
130 135 140
Met Met Val Leu Thr Leu Gly Gly Ala Phe Val Ala Tyr Ile Tyr Ser
145 150 155 160
30 Ala Pro Pro Leu Lys Leu Lys Gln Asn Gly Trp Leu Gly Asn Tyr Ala
165 170 175
Leu Gly Ala Ser Tyr Ile Ala Leu Pro Trp Trp Ala Gly His Ala Leu
180 185 190
Phe Gly Thr Leu Asn Pro Thr Ile Met Val Leu Thr Leu Ile Tyr Ser
35 195 200 205
Leu Ala Gly Leu Gly Ile Ala Val Val Asn Asp Phe Lys Ser Val Glu
210 215 220
Gly Asp Arg Gln Leu Gly Leu Lys Ser Leu Pro Val Met Phe Gly Ile
225 230 235 240
40 Gly Thr Ala Ala Trp Ile Cys Val Ile Met Ile Asp Val Phe Gln Ala
245 250 255

Gly Ile Ala Gly Tyr Leu Ile Tyr Val His Gln Gln Leu Tyr Ala Thr
 260 265 270
 Ile Val Leu Leu Leu Leu Ile Pro Gln Ile Thr Phe Gln Asp Met Tyr
 275 280 285
 5 Phe Leu Arg Asn Pro Leu Glu Asn Asp Val Lys Tyr Gln Ala Ser Ala
 290 295 300
 Gln Pro Phe Leu Val Phe Gly Met Leu Ala Thr Gly Leu Ala Leu Gly
 305 310 315 320
 His Ala Gly Ile
 10
 <210> 35
 <211> 307
 <212> PRT
 5 <213> Synechocystis sp
 <400> 35
 Met Thr Glu Ser Ser Pro Leu Ala Pro Ser Thr Ala Pro Ala Thr Arg
 1 5 10 15
 20 Lys Leu Trp Leu Ala Ala Ile Lys Pro Pro Met Tyr Thr Val Ala Val
 20 25 30
 Val Pro Ile Thr Val Gly Ser Ala Val Ala Tyr Gly Leu Thr Gly Gln
 35 40 45
 Trp His Gly Asp Val Phe Thr Ile Phe Leu Leu Ser Ala Ile Ala Ile
 50 55 60
 Ile Ala Trp Ile Asn Leu Ser Asn Asp Val Phe Asp Ser Asp Thr Gly
 65 70 75 80
 Ile Asp Val Arg Lys Ala His Ser Val Val Asn Leu Thr Gly Asn Arg
 85 90 95
 30 Asn Leu Val Phe Leu Ile Ser Asn Phe Phe Leu Leu Ala Gly Val Leu
 100 105 110
 Gly Leu Met Ser Met Ser Trp Arg Ala Gln Asp Trp Thr Val Leu Glu
 115 120 125
 Leu Ile Gly Val Ala Ile Phe Leu Gly Tyr Thr Tyr Gln Gly Pro Pro
 35 130 135 140
 Phe Arg Leu Gly Tyr Leu Gly Leu Gly Glu Leu Ile Cys Leu Ile Thr
 145 150 155 160
 Phe Gly Pro Leu Ala Ile Ala Ala Ala Tyr Tyr Ser Gln Ser Gln Ser
 165 170 175
 40 Phe Ser Trp Asn Leu Leu Thr Pro Ser Val Phe Val Gly Ile Ser Thr
 180 185 190

Ala Ile Ile Leu Phe Cys Ser His Phe His Gln Val Glu Asp Asp Leu
195 200 205
Ala Ala Gly Lys Lys Ser Pro Ile Val Arg Leu Gly Thr Lys Leu Gly
210 215 220
5 Ser Gln Val Leu Thr Leu Ser Val Val Ser Leu Tyr Leu Ile Thr Ala
225 230 235 240
Ile Gly Val Leu Cys His Gln Ala Pro Trp Gln Thr Leu Leu Ile Ile
245 250 255
Ala Ser Leu Pro Trp Ala Val Gln Leu Ile Arg His Val Gly Gln Tyr
10 260 265 270
His Asp Gln Pro Glu Gln Val Ser Asn Cys Lys Phe Ile Ala Val Asn
275 280 285
Leu His Phe Phe Ser Gly Met Leu Met Ala Ala Gly Tyr Gly Trp Ala
290 295 300

15 Gly Leu Gly
305

<210> 36

<211> 927

<212> DNA

<213> Synechocystis sp

<400> 36

atggcaacta tccaagcttt ttggcgcttc tcccgccccc ataccatcat tggtacaact 60
ctgagcgtct gggctgtgta tctgttaaact attctcgggg atggaaactc agttaactcc 120
cctgcttccc tggatttagt gttcggcgct ttgctggcct gcctgttggg taatgtgtac 180
attgtcggcc tcaaccaatt gtgggatgtg gacattgacc gcatcaataa gccgaatttg 240
cccctagcta acggagattt ttctatcgcc cagggccggt ggattgtggg actttgtggc 300
gttgccttct tggcgatcgc ctggggatta gggctatggc tggggctaac ggtgggcatt 360
30 agtttgatta ttggcacggc ctattcgggtg ccgccagtga ggttaaagcg cttttccctg 420
ctggcgggcc tgtgtattct gacgggtgcg ggaattgtgg ttaacttggg cttattttta 480
tttttttagaa ttggtttagg ttatccccc actttaataa ccccatctg ggttttgact 540
ttatttatct tagttttcac cgtggcgatc gccattttta aagatgtgcc agatatggaa 600
ggcgatcggc aatttaagat tcaaacttta actttgcaaa tcggcaaaca aaacgttttt 660
35 cggggaacct taattttact cactggttgt tatttagcca tggcaatctg gggcttatgg 720
gcggtatgc ctttaaatac tgctttcttg attgtttccc atttgtgctt attagcctta 780
ctctggtggc ggagtcgaga tgtacactta gaaagcaaaa ccgaaattgc tagtttttat 840
cagtttattt ggaagctatt tttottagag tacttgctgt atcccttggc tctgtggtta 900
cctaattttt ctaatactat ttttttag 927

<210> 37

<211> 308
<212> PRT
<213> Synechocystis sp

5 <400> 37
Met Ala Thr Ile Gln Ala Phe Trp Arg Phe Ser Arg Pro His Thr Ile
1 5 10 15
Ile Gly Thr Thr Leu Ser Val Trp Ala Val Tyr Leu Leu Thr Ile Leu
20 25 30
10 Gly Asp Gly Asn Ser Val Asn Ser Pro Ala Ser Leu Asp Leu Val Phe
35 40 45
Gly Ala Trp Leu Ala Cys Leu Leu Gly Asn Val Tyr Ile Val Gly Leu
50 55 60
Asn Gln Leu Trp Asp Val Asp Ile Asp Arg Ile Asn Lys Pro Asn Leu
65 70 75 80
Pro Leu Ala Asn Gly Asp Phe Ser Ile Ala Gln Gly Arg Trp Ile Val
85 90 95
Gly Leu Cys Gly Val Ala Ser Leu Ala Ile Ala Trp Gly Leu Gly Leu
100 105 110
20 Trp Leu Gly Leu Thr Val Gly Ile Ser Leu Ile Ile Gly Thr Ala Tyr
115 120 125
Ser Val Pro Pro Val Arg Leu Lys Arg Phe Ser Leu Leu Ala Ala Leu
130 135 140
Cys Ile Leu Thr Val Arg Gly Ile Val Val Asn Leu Gly Leu Phe Leu
145 150 155 160
Phe Phe Arg Ile Gly Leu Gly Tyr Pro Pro Thr Leu Ile Thr Pro Ile
165 170 175
Trp Val Leu Thr Leu Phe Ile Leu Val Phe Thr Val Ala Ile Ala Ile
180 185 190
30 Phe Lys Asp Val Pro Asp Met Glu Gly Asp Arg Gln Phe Lys Ile Gln
195 200 205
Thr Leu Thr Leu Gln Ile Gly Lys Gln Asn Val Phe Arg Gly Thr Leu
210 215 220
Ile Leu Leu Thr Gly Cys Tyr Leu Ala Met Ala Ile Trp Gly Leu Trp
225 230 235 240
35 Ala Ala Met Pro Leu Asn Thr Ala Phe Leu Ile Val Ser His Leu Cys
245 250 255
Leu Leu Ala Leu Leu Trp Trp Arg Ser Arg Asp Val His Leu Glu Ser
260 265 270
40 Lys Thr Glu Ile Ala Ser Phe Tyr Gln Phe Ile Trp Lys Leu Phe Phe
275 280 285

Leu Glu Tyr Leu Leu Tyr Pro Leu Ala Leu Trp Leu Pro Asn Phe Ser
290 295 300

Asn Thr Ile Phe

305

5

<210> 38

<211> 1092

<212> DNA

<213> Synechocystis sp

10

<400> 38

atgaaatttc cgccccacag tggttaccat tggcaaggtc aatcaccttt ctttgaaggt 60
tggtagctgc gcctgctttt gcccgaatcc ggggaaagtt ttgcttttat gtaactccatc 120
gaaaatcctg ctagcgatca tcattacggc ggcggtgctg tgcaaatttt agggccggct 180
acgaaaaaac aagaaaatca ggaagaccaa cttgtttggc ggacatttcc ctcggtaaaa 240
aaattttggg ccagtcctcg ccagtgtgcc ctagggcatt ggggaaaatg tagggataac 300
aggcaggcga aaccctact ctccgaagaa ttttttgcca cgggtcaagga aggttatcaa 360
atccatcaaa atcagcacca aggacaaatc attcatggcg atcgccattg tcgttggcag 420
ttcacgtag aaccggaagt aacttggggg agtcctaacc gatttcctcg ggctacagcg 480
ggttggcttt cttttttacc cttgtttgat cccggttggc aaattctttt agcccaaggt 540
agagcgcacg gctggctgaa atggcagagg gaacagtatg aatttgacca cgccctagtt 600
tatgccgaaa aaaattgggg tcaactcctt ccctcccgct ggttttggct ccaagcaaat 660
tattttcctg accatccagg actgagcgtc actgccgctg ggggggaacg gattgttctt 720
ggtcgccccg aagaggtagc tttaattggc ttacatcacc aaggtaattt ttacgaattt 780
ggcccgggcc atggcacagt cacttggcaa gtagctccct ggggcccgtt gcaattaaaa 840
gccagcaatg ataggtattg ggtcaagttg tccggaaaaa cagataaaaa aggcagttta 900
gtccacactc ccaccgcca gggtttacaa ctcaactgcc gagataccac taggggctat 960
ttgtatttgc aattgggatc tgtgggtcac ggctgatag tgcaagggga aacggacacc 1020
gcggggctag aagttggagg tgattggggg ttaacagagg aaaatttgag caaaaaaaca 1080
gtgccattct ga 1092

<210> 39

<211> 363

<212> PRT

<213> Synechocystis sp

<400> 39

Met Lys Phe Pro Pro His Ser Gly Tyr His Trp Gln Gly Gln Ser Pro

1

5

10

15

Phe Phe Glu Gly Trp Tyr Val Arg Leu Leu Leu Pro Gln Ser Gly Glu

20

25

30

Ser Phe Ala Phe Met Tyr Ser Ile Glu Asn Pro Ala Ser Asp His His
 35 40 45
 Tyr Gly Gly Gly Ala Val Gln Ile Leu Gly Pro Ala Thr Lys Lys Gln
 50 55 60
 5 Glu Asn Gln Glu Asp Gln Leu Val Trp Arg Thr Phe Pro Ser Val Lys
 65 70 75 80
 Lys Phe Trp Ala Ser Pro Arg Gln Phe Ala Leu Gly His Trp Gly Lys
 85 90 95
 Cys Arg Asp Asn Arg Gln Ala Lys Pro Leu Leu Ser Glu Glu Phe Phe
 10 100 105 110
 Ala Thr Val Lys Glu Gly Tyr Gln Ile His Gln Asn Gln His Gln Gly
 115 120 125
 Gln Ile Ile His Gly Asp Arg His Cys Arg Trp Gln Phe Thr Val Glu
 130 135 140
 15 Pro Glu Val Thr Trp Gly Ser Pro Asn Arg Phe Pro Arg Ala Thr Ala
 145 150 155 160
 Gly Trp Leu Ser Phe Leu Pro Leu Phe Asp Pro Gly Trp Gln Ile Leu
 165 170 175
 Leu Ala Gln Gly Arg Ala His Gly Trp Leu Lys Trp Gln Arg Glu Gln
 180 185 190
 20 Tyr Glu Phe Asp His Ala Leu Val Tyr Ala Glu Lys Asn Trp Gly His
 195 200 205
 Ser Phe Pro Ser Arg Trp Phe Trp Leu Gln Ala Asn Tyr Phe Pro Asp
 210 215 220
 25 His Pro Gly Leu Ser Val Thr Ala Ala Gly Gly Glu Arg Ile Val Leu
 225 230 235 240
 Gly Arg Pro Glu Glu Val Ala Leu Ile Gly Leu His His Gln Gly Asn
 245 250 255
 Phe Tyr Glu Phe Gly Pro Gly His Gly Thr Val Thr Trp Gln Val Ala
 260 265 270
 30 Pro Trp Gly Arg Trp Gln Leu Lys Ala Ser Asn Asp Arg Tyr Trp Val
 275 280 285
 Lys Leu Ser Gly Lys Thr Asp Lys Lys Gly Ser Leu Val His Thr Pro
 290 295 300
 35 Thr Ala Gln Gly Leu Gln Leu Asn Cys Arg Asp Thr Thr Arg Gly Tyr
 305 310 315 320
 Leu Tyr Leu Gln Leu Gly Ser Val Gly His Gly Leu Ile Val Gln Gly
 325 330 335
 Glu Thr Asp Thr Ala Gly Leu Glu Val Gly Gly Asp Trp Gly Leu Thr
 340 345 350
 40 Glu Glu Asn Leu Ser Lys Lys Thr Val Pro Phe

355

360

<210> 40

<211> 56

5 <212> DNA

<213> Artifical Sequence

<400> 40

cgcgatttaa atggcgcgcc ctgcaggcgg ccgcctgcag ggcgcgccat ttaaat 56

10

<210> 41

<211> 32

<212> DNA

<213> Artifical Sequence

<400> 41

tcgaggatcc gcggcgcaa gcttcctgca gg 32

<210> 42

<211> 32

<212> DNA

<213> Artifical Sequence

<400> 42

tcgacctgca ggaagcttgc ggccgcggat cc 32

<210> 43

<211> 32

<212> DNA

<213> Artifical Sequence

<400> 43

tcgacctgca ggaagcttgc ggccgcggat cc 32

<210> 44

<211> 32

<212> DNA

<213> Artifical Sequence

<400> 44

tcgaggatcc gcggcgcaa gcttcctgca gg 32

<210> 45
<211> 36
<212> DNA
5 <213> Artificial Sequence

<400> 45
tcgaggatcc gcggccgcaa gcttcctgca ggagct 36

10 <210> 46
<211> 28
<212> DNA
<213> Artificial Sequence

15 <400> 46
cctgcaggaa gcttgccggc gcggatcc 28

20 <210> 47
<211> 36
<212> DNA
<213> Artificial Sequence

25 <400> 47
tcgacctgca ggaagcttgc ggccgcggat ccagct 36

30 <210> 48
<211> 28
<212> DNA
<213> Artificial Sequence

35 <400> 48
ggatccgcgg ccgcaagctt cctgcagg 28

40 <210> 49
<211> 39
<212> DNA
<213> Artificial Sequence

<400> 49
gatcacctgc aggaagcttg cggccgcgga tccaatgca 39

5
10
15
20
25
30
35
40

<210> 50
<211> 31
<212> DNA
<213> Artificial Sequence

<400> 50
ttggatccgc ggccgcaagc ttcctgcagg t 31

<210> 51
<211> 41
<212> DNA
<213> Artificial Sequence

<400> 51
ggatccgcgg ccgcacaatg gagtctctgc tctctagttc t 41

<210> 52
<211> 38
<212> DNA
<213> Artificial Sequence

<400> 52
ggatcctgca ggtcaattca aaaaaggtaa cagcaagt 38

<210> 53
<211> 45
<212> DNA
<213> Artificial Sequence

<400> 53
ggatccgcgg ccgcacaatg gcgttttttg ggctctcccg tgttt 45

<210> 54
<211> 40
<212> DNA
<213> Artificial Sequence

<400> 54
ggatcctgca ggttattgaa aacttcttcc aagtacaact 40

<210> 55

<211> 38
<212> DNA
<213> Artificial Sequence

5 <400> 55
ggatccgcgg ccgcacaatg tggcgaagat ctgttgtt 38

<210> 56
<211> 37

10 <212> DNA
<213> Artificial Sequence

<400> 56
ggatcctgca ggtcatggag agtagaagga aggagct 37

<210> 57
<211> 50
<212> DNA
<213> Artificial Sequence

<400> 57
ggatccgcgg ccgcacaatg gtacttgccg aggttccaaa gcttgccctct 50

<210> 58
<211> 38
<212> DNA
<213> Artificial Sequence

30 <400> 58
ggatcctgca ggtcacttgt ttctggtgat gactctat 38

<210> 59
<211> 38
<212> DNA

35 <213> Artificial Sequence

<400> 59
ggatccgcgg ccgcacaatg acttcgattc tcaacact 38

40 <210> 60
<211> 36

<212> DNA

<213> Artifical Sequence

<400> 60

5 ggatcctgca ggtcagtggt gcgatgctaa tgccgt 36

<210> 61

<211> 22

<212> DNA

10 <213> Artifical Sequence

<400> 61

taatgtgtac attgtcggcc tc 22

15 <210> 62

<211> 60

<212> DNA

<213> Artifical Sequence

20 <400> 62

gcaatgtaac atcagagatt ttgagacaca acgtggcttt ccacaattcc ccgcaccgtc 60

<210> 63

<211> 22

25 <212> DNA

<213> Artifical Sequence

<400> 63

30 aggctaataa gcacaaatgg ga 22

<210> 64

<211> 63

<212> DNA

35 <213> Artifical Sequence

<400> 64

ggtatgagtc agcaacacct tcttcacgag gcagacctca gcggaattgg tttaggttat 60

ccc 63

40 <210> 65

<211> 26

<212> DNA

<213> Artifical Sequence

<400> 65

5 ggatccatgg ttgcccaaac cccatc 26

<210> 66

<211> 61

<212> DNA

10 <213> Artifical Sequence

<400> 66

gcaatgtaac atcagagatt ttgagacaca acgtggcttt gggtaagcaa caatgaccgg 60
c 61

<210> 67

<211> 25

<212> DNA

<213> Artifical Sequence

<400> 67

gaattctcaa agccagccca gtaac 25

<210> 68

<211> 63

<212> DNA

<213> Artifical Sequence

<400> 68

30 ggtatgagtc agcaacacct tcttcacgag gcagacctca gcggtgcga aaagggtttt 60
ccc 63

<210> 69

<211> 23

35 <212> DNA

<213> Artifical Sequence

<400> 69

ccagtggttt aggctgtgtg gtc 23

40

<210> 70

<211> 21
<212> DNA
<213> Artificial Sequence

5 <400> 70
ctgagttgga tgtattggat c 21

<210> 71
<211> 28
10 <212> DNA
<213> Artificial Sequence

<400> 71
ggatccatgg ttacttcgac aaaaatcc 28

15 <210> 72
<211> 60
<212> DNA
<213> Artificial Sequence

20 <400> 72
gcaatgtaac atcagagatt ttgagacaca acgtggcttt gctaggcaac cgcttagtac 60

25 <210> 73
<211> 28
<212> DNA
<213> Artificial Sequence

30 <400> 73
gaattcttaa cccaacagta aagttccc 28

<210> 74
<211> 63
<212> DNA
35 <213> Artificial Sequence

<400> 74
ggtatgagtc agcaacacct tcttcacgag gcagacotca gcgcggcat tgtcttttac 60
atg 63

40 <210> 75

5 <211> 20
<212> DNA
<213> Artificial Sequence

5 <400> 75
ggaacccttg cagccgcttc 20

<210> 76
<211> 22
10 <212> DNA
<213> Artificial Sequence

<400> 76
gtatgcccaa ctggtgcaga gg 22

15 <210> 77
<211> 28
<212> DNA
<213> Artificial Sequence

20 <400> 77
ggatccatgt ctgacacaca aaataccg 28

25 <210> 78
<211> 62
<212> DNA
<213> Artificial Sequence

30 <400> 78
gcaatgtaac atcagagatt ttgagacaca acgtggcttt cgccaatacc agccaccaac 60
ag 62

35 <210> 79
<211> 27
<212> DNA
<213> Artificial Sequence

40 <400> 79
gaattctcaa atccccgcat ggccctag 27

<210> 80

<211> 65
<212> DNA
<213> Artificial Sequence

5 <400> 80
gggatgagtc agcaacacct tcttcacgag gcagacctca gcggcctacg gcttggacgt 60
gtggg 65

<210> 81
10 <211> 21
<212> DNA
<213> Artificial Sequence

<400> 81
15 cacttggatt cccctgatct g 21

<210> 82
<211> 21
<212> DNA
20 <213> Artificial Sequence

<400> 82
gcaatacccg cttggaaaac g 21

<210> 83
25 <211> 29
<212> DNA
<213> Artificial Sequence

30 <400> 83
ggatccatga ccgaatcttc gccoctage 29

<210> 84
<211> 61
35 <212> DNA
<213> Artificial Sequence

<400> 84
gcaatgtaac atcagagatt ttgagacaca acgtggcttt caatcctagg tagccgaggc 60
40 g 61

5
10
15
20
25
30
35
40

<210> 85
<211> 27
<212> DNA
<213> Artifical Sequence

<400> 85
gaattcttag cccaggccag cccagcc 27

<210> 86
<211> 66
<212> DNA
<213> Artifical Sequence

<400> 86
ggtatgagtc agcaacacct tcttcacgag gcagacctca gcggggaatt gatttgttta 60
attacc 66

<210> 87
<211> 21
<212> DNA
<213> Artifical Sequence

<400> 87
gcgatcgcca ttatcgcttg g 21

<210> 88
<211> 24
<212> DNA
<213> Artifical Sequence

<400> 88
gcagactggc aattatcagt aacg 24

<210> 89
<211> 25
<212> DNA
<213> Artifical Sequence

<400> 89
ccatggattc gagtaaagtt gtcgc 25

<210> 90
<211> 0
<213> Artificial Sequence

5 <400> 90
gaattcactt caaaaaaggt aacag

<210> 91
<211> 4550

10 <212> DNA
<213> Arabidopsis sp

<400> 91

15	attttacacc aatttgatca cttaactaaa ttaattaaat tagatgatta tcccaccata	60
	tttttgagca ttaaaccata aaaccatagt tataagtaac tgttttaatac gaatatgact	120
	cgattaagat taggaaaaat ttataaccgg taattaagaa aacattaacc gtagtaaccg	180
	taaatgccga ttcttccctt gtctaaaaga cagaaaacat atattttatt ttgccccata	240
	tgtttcactc tatttaattt caggcacaat acttttggtt ggtaacaaaa ctaaaaagga	300
	caacacgtga tacttttcct cgtccgtcag tcagattttt tttaaactag aaacaagtgg	360
20	caaactctaca ccacattttt tgcttaatct attaacttgt aagtttttaa ttcctaaaaa	420
	agtctaacta attcttctaa tataagtaca ttccctaaat ttccccaaaa gtcaaattaa	480
	taattttcaa aatctaactt aaatatctaa taattcaaaa tcattaaaaa gacacgcaac	540
	aatgacacca attaatcatc ctgcaccac acaattctac agttctcatg ctaaaccata	600
	ttttttgctc tctgttcctt caaaatcatt tctttctctt ctttgattcc caaagatcac	660
25	ttctttgtct ttgatttttg attttttttc tctctggcgt gaaggaagaa gctttatttc	720
	atggagtctc tgctctctag ttcttctctt gtttcogctg gttaaactctg tccttttctg	780
	gtttcagggt ttatttggtg tttaggtttc gtttttgtga ttcagaacca tacaaaaagt	840
	ttgaactttt ctgaatataa aataaggaaa aagtttcgat ttttataatg aattgtttac	900
	tagatcgaag taggtgacaa aggttattgt gtggagaagc ataatttctg ggcttgactt	960
30	tgaattttgt ttctcatgca tgcaacttat caatcagctg gtgggttttg ttggaagaag	1020
	cagaatctaa agctccactc tttatcaggt tcgttagggt tttatgggtt tttgaaatta	1080
	aatactcaat catcttagtc tcattattct attgggttgaa tcacattttc taatttgga	1140
	tttatgagac aatgtatgtt ggacttagtt gaagttcttc tctttgggta tagttgaagt	1200
	gttactgatg ttgttttagct ctttacacca atatatacac ccaattttgc agaaatccga	1260
35	gttctgcgtt gtgattcgag taaagttgtc gcaaaaccga agtttaggaa caatcttggt	1320
	aggcctgatg gtcaaggatc ttcattggtt ttgtatccaa aacataagtc gagatttcgg	1380
	gttaatgcc aatgcgggtc gcttgaggct ttcgactcga atagcaaaca gaagtctttt	1440
	agagactcgt tagatgcgtt ttacagggtt tctaggcctc atacagttat tggcacagtt	1500
	aagtttctct ttaaaaatgt aactctttta aaacgcaatc tttcagggtt ttcaaggaga	1560
40	taacattagc tctgtgattg gatttgcagg tgcttagcat tttatctgta tctttcttag	1620
	cagtagagaa ggtttctgat atatctcctt tacttttcac tggcatcttg gaggtaatga	1680

atataataaca cataatgacc gatgaagaag atacattttt ttcgtctctc tgttttaaca 1740
attgggtttt gttttcaggc tgttggtgca gctctcatga tgaacattta catagttggg 1800
ctaaatcagt tgtctgatgt tgaaatagat aaggtaacat gcaaattttc ttcatatgag 1860
ttcgagagac tgatgagatt aatagcagct agtgcctaga tcatctctat gtgggttttt 1920
5 gcaggttaac aagccctatc ttccattggc atcaggagaa tattctgtta acaccggcat 1980
tgcaatagta gcttccttct ccatcatggg atgggtgccat tttcacaaaa tttcaacttt 2040
tagaattcta taagttactg aaatagtttg ttataaatcg ttatagagtt tctgggttgg 2100
gtggattgtt ggttcattggc cattgttctg ggctcttttt gtgagtttca tgctcggtag 2160
tgcatactct atcaatgtaa gtaagtttct caataactaga atttggtctca aatcaaaatc 2220
10 tgcagtttct agtttttaggt taatgaggtt ttaataactt acttctacta caaacagttg 2280
ccacttttac ggtggaaaag atttgcatg gttgcagcaa tgtgtatcct cgctgtccga 2340
gctattattg ttcaaatcgc cttttatcta catattcagg tactaaacca ttttccttat 2400
gttttgtagt tgttttcac aaaatcactt ttatattact aaagctgtga aactttgttg 2460
cagacacatg tgtttggaag accaatcttg ttcactaggc ctcttatttt cgccactgcg 2520
5 tttatgagct ttttctctgt cgttattgca ttgtttaagg taaacaaaga tggaaaaaga 2580
ttaaatctat gtatacttaa agtaaagcat tctactgtta ttgatgagaa gttttctttt 2640
ttggttggtg gcaggatata cctgatatcg aaggggataa gatattcgga atccgatcat 2700
tctctgtaac tctgggtcag aaacgggtac gatattctaa cttaaagaaat tgttttgact 2760
caagtgttgg attaaagatta cagaagaaag aaaactgttt ttgtttcttg caaaattcag 2820
20 gtgttttgga catgtgttac actacttcaa atggcttaag ctgttgcaat tctagttgga 2880
gccacatctc cattcatatg gagcaaagtc atctcggtaa caatctttct ttacccatcg 2940
aaaactcgct aattcatcgt ttgagtggtg ctgggttccat tttgttccgt tctgttgatt 3000
ttttttcagg ttgtgggtca tgttatactc gcaacaactt tgtgggctcg agctaagtcc 3060
gttgatctga gtagcaaaac cgaaataact tcatgttata tgttcatatg gaaggttaga 3120
5 ttcgtttata aatagagtct ttactgcctt tttatgcgct ccaatttgga attaaaatag 3180
cctttcagtt tcatcgaatc accattatac tgataaatc tcatctctgc atcagctctt 3240
ttatgcagag tacttgctgt tacctttttt gaagtgaact acattagaag agaagaagat 3300
ggagataaaa gaataagtca tccatctgct tctgttttta ttacaagttc atgaaattag 3360
gtagtgaact agtgaattag agttttatct tgaacatgg cagactgcaa aaatatgtca 3420
30 aagatatgaa tttctgttgg gtaaagaagt ctctgcttgg gcaaaatctt aaggttcggt 3480
gtgttgatat aatgctaagc gaagaaatcg attctatgta gaaatttccg aaactatgtg 3540
taaacatgtc agaacatctc cattctatat cttcttctgc aagaaagctc tgtttttatc 3600
acctaaactc tttatctctg tgtagttaag atatgtatat gtacgtgact acattttttt 3660
gttgatgtaa tttgcagaac gtatggattt ttgttagaaa gcatgagttc gaaagtatat 3720
35 gtttatatat atggataatt cagacctaac gtcgaagctc acaagcataa attcactact 3780
atagtttgct ctgtaataga tagttccatt gatgtcttga aactgtacgt aactgcctgg 3840
gcgttttgtg gttgatactg actactgagt gttctttgtg agtgttgtaa gtatacaaga 3900
agaagaatat aggctcacgg gaacgactgt ggtggaagat gaaatggaga tcatcacgta 3960
goggctttgc caaagaccga gtcacgatcg agtctatgaa gtctttacag ctgctgatta 4020
40 tgattgacca ttgcttagag acgcattgga atcttactag ggacttgcct gggagtttct 4080
tcaagtacgt gtcagatcat acgatgtagg agatttcacg gctttgatgt gtttggttgg 4140

agtcacaatg cttaatgggc ttattggccc aataatagct agctcttttg ctttagccgt 4200
 ttogttttgtc ccttggtggt gagtattatt agggtagtgt gtgaccaaag tcaccagacc 4260
 tagagtgaat ctagtagagt cctagaccat ggtccatggc ttttatttgt aatttgaaaa 4320
 atgaacaatt ctttttgtaa ggaaaacttt tatatagtag acgtttacta tatagaaact 4380
 5 agttgaacta acttcgtgca attgcataat aatgggtgtga aatagagggt gcaaaactca 4440
 ataaacatth cgacgtacca agagttcgaa acaataagca aaatagattt ttttgcttca 4500
 gactaatttg tacaatgaat ggtaataaaa ccattgaagc ttttattaat 4550

<210> 92
 10 <211> 4450
 <212> DNA
 <213> Arabidopsis sp

<400> 92

15 tttaggttac aaaatcaatg atattgcgta tgtcaactat aaaagccaaa agtaaagcct 60
 cttgtttgac cagaagggtca tgatcattgt atacatacag ccaaactacc tcctggaaga 120
 aaagacatgg atcccaaaca acaacaatag cttcttttac aagaaccagt agtaactagt 180
 cactaatcta aaagagttaa gtttcagctt ttctggcaat ggctccttga tcatttcaat 240
 cctgaaggag acccactttg tagcaagacc atgtcctctg tttcacttac agtgtgtctc 300
 20 aaaagtctac ttcaattctt catatatagg ttcttcacac tacagcttca tcctcattcg 360
 ttgacagaga gagagtcttt attgaaaact tcttccaagt acaactccac taaatataat 420
 agcaccaaac cacttggtcg acacaaatct gtacagatat aaaaactata ttaggttttc 480
 caaggcaaat cacataattg gattgtgaaa gagtacaaaa gataaaccca aattttcata 540
 ctttctactg cagtcagcac cagatgataa gtcagctgtc cctatttgcc atcctaactg 600
 25 tcctgatgca ggggccagtg atggcgaata ttgccacct taatcattag agcgagaaac 660
 aaaaagaatc aaaagacagt aaatggaatt aggaatcaca aatgagtcct tgtaaagtth 720
 attgagtacc gagatctgca ctgaatccag aaagtgcaag aaaacctatg gatgctgtgc 780
 caaatccagt taaccaaagc tttgtattat caccgaatct aagggtctgt gacttaacac 840
 caactttttac atcatcttct ttgtcctgga gacacaatat attagacatt agtccatgga 900
 30 aaaaaaatga tttaacctag aatatctcaa aattacttgc ataaaaactg aacttgagct 960
 gaaatthtg gttcgtagct tgtggcatat actatttcat tttcaatggg ccacaaaggt 1020
 aactttcttt tctcacttct gttgcaaacg ggaagactth tatggggcta actcttcact 1080
 taaagtatag aaatcagatg gaaaagggtg gagatcaggg taattttctt ctttatgatt 1140
 gacaaaagtc gaacatcgaa atggatgcat ttgcatgaga catgaaacaa aagctgaaaa 1200
 35 agaaatctgt ggtggtgaag ctagaaaaag aaaacaaagc aagcaatatg cacacattga 1260
 gattaactac tttgctactg gtcataatca aatagattth gaagctaaaa aataaaaagt 1320
 gaatatacct gatgtgcata aatagtatca taaacaaggg tccagcagac tccggagaga 1380
 tagagagggg gtacaataga tgggtctatg cttcctthaa ctgcagtcca tcttaacaat 1440
 gctccccagt ttatggtcaa acctaaaaag gcttgaggct gcaattataa aaacgaatca 1500
 40 atcataagaa aatcagaaaa tatataatgt ctaactthga gaagccagaa tagattthaa 1560
 ttacccaaaa tgtaaacctc ttcataagtg ggtaggaaaa gacaagtaac aaagatgaag 1620

	ccccataaac	acggctgcag	aatatacata	ctgaaatgag	ctcaagtaga	aaagaatttg	1680
	atcacaaaaac	taaagacaag	acctgagaac	atatcttcag	aattttgggc	aactacataa	1740
	gggtgaacca	tatgtgtatg	tgaattttta	aacaaacact	tgcaaatacg	cgacttttagg	1800
	gcaagtaaaa	aatccaaaca	aacctgtaat	tgttaagttg	gagaagaatc	cctaagccta	1860
5	aaagcaactg	cagcccgaga	aatccaatcc	cttgaaatgg	tgtcaaaaaga	ccactggcga	1920
	taggtottag	ttttgtacga	tcaacctgga	tataaaaagaa	atttgtaaga	caacataatc	1980
	taaaacaaaa	caaccatata	aaatcttgag	ctttacatac	aagcaaccca	tottttgttta	2040
	tggaagaatg	aatccagtta	catgaatgct	gtgtatctac	cctaactact	aaacacatat	2100
	ttcaatcgaa	aaacatatte	caccttcacc	atatctaaca	cctgaagtct	ttcacttttt	2160
10	gaacgaagtc	atcagaacat	gcagataagc	tattacccaa	aacagagata	tgactggaaa	2220
	tgttgtcgta	aattgatcca	acatagaaaa	atcaagacca	gttccagatg	tcaaagcaat	2280
	aacactttcc	caccatgggt	acagaaacca	tagttacaca	aaacatgttt	cctaaaccaa	2340
	catactaaag	ggatatataa	atttgacatc	actttatcac	cataccataa	gatagcttaa	2400
	aaacaaactg	acctttgtat	ctatgtcctg	atcaagcaga	tcattttatag	tacaaccagc	2460
15	acctctaaga	agtaatgctc	cgcaacccaa	taaagccata	tattttaaacc	ttggaaggct	2520
	tccaggatca	gcagccaacg	caatcgacct	atacaacaat	gatggagatt	cagagtatcg	2580
	atctattttac	atagctctgg	aactagatcc	atgacgaaac	atggaacatc	gttataatat	2640
	ctaaagactt	ccaaacagat	tcctgagtaa	gaaacccagt	ggaactatag	tactgtaaca	2700
	tatataaaat	caaagaaaac	tcaggtttat	agcattatcc	aatcctgatt	tctgccaatc	2760
20	cttaaccact	ctcccatgct	atcaaaaacc	tcagctcaag	atcatactac	ctaattgcct	2820
	atgagctctt	gggaagatca	ttatggattt	gataactgaa	aaaagtaaca	gagaaatagc	2880
	agactgcaag	aactactcca	aacttctcca	ctgatatgta	tgtagtctaa	caataataaa	2940
	cagacataaa	ttotttttatc	aagcttcaag	agcaagttag	tcagaaaaca	tcacagccaa	3000
	accaaccagg	aaaacacata	acttttatcac	ataaaaactaa	atttaaatgta	atctgactta	3060
25	acataaacca	tccttttgga	cgaaaggaaa	ctatataaac	atgcagtctt	tctttccctc	3120
	agctattctt	toggatggat	tataatgaat	ctcaaaagtg	aaatgtcttg	attctcagct	3180
	acattactca	aaggcgaaga	taaacttacc	acatacaagg	ccacgcaagc	aaccaagttc	3240
	caatgggttt	atccaatoga	gcaagcttag	cataacctct	aacttcttct	ggtaaataca	3300
	aatctatcca	agaagcttcc	ttaacaacaa	caccatcact	cttctcctta	tcctctttct	3360
30	toggctttcc	ctccaaaacc	gaagaagacg	acgacattcc	acaaattaat	ctgtaattcc	3420
	aaccaacacc	aaaaaacttc	tcctgatgca	attctcttcc	tttactccat	acttggtaat	3480
	tatcattcca	tgaaggataa	cacttagtga	aaggatttgt	gtaatgggta	gtcacaggat	3540
	tggacaagga	tttatgttgt	gattgcaaaa	gagcagagga	agaagatgga	gttacggaga	3600
	cggaagattt	caacaaccgt	cttgaaacac	gggagagccc	aaaaaacgcc	atctttgaga	3660
35	gaaattgttg	cctggaagaa	acaaagactt	gagatttcaa	acgtaagtga	attctttacga	3720
	acgaaagcta	acttctcaag	agaatcagat	tagtgattcc	tcaaaaacaa	acaaaactat	3780
	ctaatttcag	tttcgagtga	tgaagcctta	agaatctaga	acctccatgg	cgtttctaata	3840
	ctctcagaga	taatcgaatt	ccttaaacaa	tcaaagctta	gaaagagaag	aacaacaaca	3900
	acaacaaaaa	aatcagatt	aacaaccgac	cagagagcaa	cgacgacgcc	ggcgagaaag	3960
40	agcacgtcgt	ctcgagcaa	gacttcttct	ccagtaaccc	ggatggatcg	ttaatgggcc	4020
	tgtagattat	tatatttggg	ccgaaacaat	tgggtcagca	aaaacttggg	ggataatgaa	4080

gaaacacgta cagtatgcat ttaggctcca aattaattgg ccatataatt cgaatcagat 4140
aaactaatca acccctacct tacttatttc tcaactgtttt tatttctacc ttagtagttg 4200
aagaaacact tttattttatc ttttcgggac ccaaatttga taggatcggg ccattactca 4260
tgagcgtcag acacatatta gccttatcag attagtgggg taagggtttt ttaattcggg 4320
5 aagaagcaac aatcaatgtc ggagaaatta aagaatctgc atgggcgtgg cgtgatgata 4380
tgtgcatatg gagtcagttg ccgatcatat ataactattt ataaactaca tataaagact 4440
actaatagat 4450

<210> 93
10 <211> 2850
<212> DNA
<213> Arabidopsis sp

<400> 93
15 aattaaaatt tgagcggctc aaaccattag accgtttaga gatccctcca acccaaaata 60
gtcgattttc acgtcttgaa catatattgg gccttaatct gtgtgggttag taaagacttt 120
tattggtcaa agaaaaacaa ccatggccca acatgttgat acttttattt aattatacaa 180
gtaccctga attctotgaa atatatattga ttgaccaga tattaatttt aattatcatt 240
tctgtaaaa gtgaaggagt cacogtgact cgtcgtaato tgaaaccaat ctgttcatat 300
20 gatgaagaag tttctctcgt tctcctccaa cgcgtagaaa attctgacgg cttacgatg 360
tggcgaagat ctgttggtta tegtctctct tcaagaatct ctgtttcttc ttcgttacca 420
aaccctagac tgattccttg gtcccgcgaa ttatgtgccg ttaatagctt ctccagcct 480
cgggtctoga cggaaatcaac tgctaagtta gggatcactg gtgttagatc tgatgccaat 540
cgagtttttg ccaactgctac tgccgcgcgt acagctacag ctaccaccgg tgagatttcg 600
25 tctagagttg cggctttggc tggattaggg catcactacg ctcgttgtta ttgggagctt 660
tctaaagcta aacttaggta tgtgtttact tttcttttct catgaaaaat ctgaaaattt 720
ccaattggtg gattcttaaa ttctcatttg ttttatgggt gtagtatgct tgtggttgca 780
acttctggaa ctgggtatat tctgggtacg ggaaatgctg caattagctt cccggggctt 840
tgttacacat gtgcaggaac catgatgatt gctgcatctg ctaattcctt gaatcaggtc 900
30 attgaaatgt tgagaagttc ataaatttcg aatccttggt gtgtttatgt agttgatctt 960
gottgcttat gtttatgtag ttgaaaagtt taaaaatttc taatccttg tagttgatct 1020
cgcttggttg ttttttcatt ttctagattt ttgagataag caatgattct aagatgaaaa 1080
gaacgatgct aaggccattg ccttcaggac gtattagtgt tccacacgct gttgcatggg 1140
ctactattgc tgggtcttct ggtgcttggt tgttgccag caaggtgaat gtttggtttt 1200
35 ttatatgtga tttctttggt ttatgaatgg gtgattgaga gattatggat ctaaaatttt 1260
gcttccacga caaggttatt gcagactaat atgttggtg ctggacttgc atctgccaat 1320
cttgtacttt atgcgtttgt ttatactccg ttgaagcaac ttcacctat caatacatgg 1380
gttggcgctg ttgttggtgc tatccaccc ttgcttggtt aaatttttgt tcttttctt 1440
ctttatttta gcagattctg ttttggttga tactgctttt aattcaaaat gtagtcatgg 1500
40 ttcaccaatt ctatgcttat ctattttgtg tgttgctcag tgggcggcag cgtctggtca 1560
gatttcatac aattcgatga ttcttcacgc tgctctttac ttttggcaga tacctcattt 1620

5
10
15
20
25
30
35
40

tatggccctt gcacatctct gccgcaatga ttatgcagct ggagggttaag accatatggt 1680
gtcatatgag attagaatgt ctcttccat gtagtggtga tcttgaacta gttcaatttc 1740
gtggaatgat cagagtgtcc tagatagtgt cacagcagtc gacatttttag tggctagata 1800
atgagttctt tccgttagag ataaacattc gcgaacattg tttccagctt ccgcgaccca 1860
acttctgatt ttgtttcttg gtacctgtt ttcagttaca agatgttgtc actctttgat 1920
ccgtcagggg agagaatagc agcagtggct ctaaggaaact gctttttacat gatccctctc 1980
ggtttcatcg cctatgactg tgagtcttgt agattcatct tttttttgta gtttattgac 2040
tgcattgctg tatctgattt ttgctgttcc ttccaatttt tgtgacaggg gggttaacct 2100
caagttgggt ttgctctgaa tcaacacttc tcacactagc aatcgctgca acagcatttt 2160
cattctaccg agaccggacc atgcataaag caaggaaaat gttccatgcc agtcttctct 2220
tccttctctg tttcatgtct ggtcttcttc tacaccgtgt ctctaattgat aatcagcaac 2280
aactcgtaga agaagccgga ttaacaaatt ctgtatctgg tgaagtcaaa actcagaggc 2340
gaaagaaacg tgtggctcaa cctccggtgg cttatgcctc tgcctgcaccg tttcctttcc 2400
tccagctcc ttccttctac tctccatgat aacctttaag caagctattg aatttttgga 2460
aacagaaatt aaaaaaaaaa tctgaaaagt tcttaagttt aatctttggg taataatgaa 2520
gtggagaacg catacaagtt tatgtatttt ttctcatctc cacataattg tattttttct 2580
ctaagtatgt ttcaaatgat acaaaaataca tactttatca attatctgat caaattgatg 2640
aatttttgag ctttgacgtg ttaggtctat ctaataaaac tagtaacgaa tttggttttg 2700
gaaatgaaat ccgataaccg atgatgggtg agagttaaac gattaaaccg ggttggttaa 2760
aggtctcgag tctcgacggc tgcggaaatc ggaaaatcac gattgaggac tttgagctgc 2820
cacgaagatg gcgatgaggt tgaaatcaat 2850

<210> 94
<211> 3660
<212> DNA
<213> Arabidopsis sp

<400> 94

30
35
40

tatttgtatt ttattgtta aattttatga tttcaccgg tatatatcat cccatattaa 60
tattagattt attttttggg ctttatttg gttttcgatt taaactgggc ccattctgct 120
tcaatgaaac cctaattgggt ttgttttggg ctttgattt aaaccgggcc cattctgctt 180
caatgaaggt cttttgtcca acaaaactaa catccgacac aactagtatt gccagagga 240
tcgtgccaca tggcagttat tgaatcaaag gccgccaaaa ctgtaacgta gacattactt 300
atctccggta acggacaacc actcgtttcc cgaaacagca actcacagac tcacaccact 360
ccagtctcog gcttaactac caccagagac gattctctct tccgtcgggt ctatgaactc 420
gattctcaac actgtctcca ccatccactc ttccagagtt acctccgtcg atcgagtcgg 480
agtcctctct cttegggaatt cggattccgt tgagttcact cgccggcggt ctggtttctc 540
gacgttgatc tacgaatcac ccggtagtta gcattctgtt ggatagattg atgaatgttt 600
tcttcgattt tttttttact gatcttggtg tggatctctc gtagggcgga gatttggtgt 660
gcgtgcccgc gagactgata ctgataaagg tatgattttt tagttgtttt tattttctct 720
ctcttcaaaa ttctcttttc aaacactgtg gcgtttgaat ttccgacggc agttaaatct 780

5
10
15
20
25
30
35
40

cagacacctg acaaggcacc agccgggtgt tcaagcatta accagcttct cggtatcaaa 840
ggagcatctc aagaaactgt aattttgttc atctcctcag aatcttttaa attatcatat 900
ttgtggataa tgatgtgtta gtttaggaat ttctacta aaggtaatct cttttgagga 960
caagtcttgt ttttagctta gaaatgatgt gaaaatgttg tttgttagct aaaaagagtt 1020
5 tgttgttata ttctgtatct agaataaatg gaagattcgt cttcagctta caaaaccagt 1080
cacttggcct ccactgggtt ggggagtcgt ctgtggtgct gctgcttcag gtaatcatac 1140
gaacctcttt tggatcatgc aatactgtac agaaagtttt ttcatTTTTcc ttccaattgt 1200
ttcttctggc agggaacttt cattggaccc cagaggatgt tgctaagtcg attctttgca 1260
tgatgatgtc tggctcttgt cttactggct atacacaggt ctggttttac acaacaaaaa 1320
10 gctgacttgt tcttattcta gtgcatttgc ttggtgctac aataacctag acttgtcgat 1380
ttccagacaa tcaacgactg gtatgataga gatatcgacg caattaatga gccatatcgt 1440
ccaattccat ctggagcaat atcagagcca gaggtaactg agacagaaca ttgtgagctt 1500
ttatctcttt tgtgattctg atttctcctt actccttaaa atgcagggtta ttacacaagt 1560
ctgggtgcta ttattgggag gtcttggtat tgctggaata ttagatgtgt gggtaagttg 1620
15 gcccttctga cattaactag tacagttaaa gggcacatca gatttgctaa aatcttccct 1680
tatcaggcag ggcataccac tccactgtc ttctatcttg ctttgggagg atcattgcta 1740
tcttatatat actctgtctc acctcttaag gtaagtttta ttcttaactt ccactctcta 1800
gtgataagac actccatcca agttttggag ttttgaatat cgatatctga actgatctca 1860
ttgcagctaa aacaaaatgg atgggttga aattttgcac ttggagcaag ctatattagt 1920
20 ttgccatggt aagatatctc gtgtatcaat aatatatggc gttgttctca tctcattgat 1980
ttgtttcttg ctcaattgac tgataggttg gctggccaag cattgttttg cactcttacg 2040
ccagatgttg ttgttctaac actcttgtag agcatagctg gggtagctct ttggcaaacc 2100
ttttatgttg cttttttcgt tatctgttgt aatatgtctt tgcttcatgt tgtaccttg 2160
tgataatgca gtttaggaata gccattgtta acgacttcaa aagtgttgaa ggagatagag 2220
25 cattaggact tcagtctctc ccagtagctt ttggcacoga aactgcaaaa tggatatgag 2280
ttggtgctat agacattact cagctttctg ttgcccgtat gtactatcca ctgtttttgt 2340
gcagctgttg cttctatttc ttttcttga tcttatcaac tggatattca ccaatggtaa 2400
agcacaatt aatgaagctg aatcaacaaa ggcaaacat aaaagtacat tctaataaaa 2460
tgagctaata aagaggaggc atctactttt atgtttcatt agtgtgattg atggattttc 2520
30 atttcatgct tctaaaacaa gtattttcaa cagtgtcatg aaataacaga acttatatct 2580
tcatttgtac ttttactagt ggatgagttt cacaatcatt gttatagaac caaatcaaa 2640
gtagagatca tcattagtat atgtctatct tgggtgagc atactctatta gcatctggga 2700
aaccttatta tgcgttggcg ttggttgcct tgatcattcc tcagattgtg ttccaggtaa 2760
agacgttaac agtctcacat tataattaat caaattcttg tcaactcgtc gattgctaca 2820
35 ctgccttcta taaactgcag tttaaatact ttctcaagga ccctgtcaaa tacgacgtca 2880
agtaccaggt aagtcaactt agtacacatg tttgtgttct tttgaaatat ctttgagagg 2940
tctcttaata agaagttgct tgaaacactc atcttgatta caggcaagcg cgcagccatt 3000
cttgggtgct ggaatatttg taacggcatt agcatcgcaa cactgaaaaa ggcgtatttt 3060
gatgggggtt tgtcgaaagc agaggtgttg acacatcaaa tgtgggcaag tgatggcatc 3120
40 aactagttta aaagattttg taaaatgtat gtaccgttat tactagaaac aactcctgtt 3180
gtatcaattt agcaaacagg ctgagaaatt gtaattgatg ttaccgtatt tgcgtccat 3240

ttttgcattt cctgctcata tcgaggattg gggtttatgt tagttctgtc acttctctgc 3300
tttcagaatg tttttgtttt ctgtagtgga ttttaactat tttcatcact ttttgtattg 3360
attctaaaca tgtatccaca taaaaacagt aatatacaaa aatgatactt cctcaaactt 3420
tttataatct aaatctaaca actagctagt aoccoaacta acttcatata attaatattga 3480
5 gaaactacaa agactagact atacatatgt tatttaacaa cttgaaactg tgttattact 3540
acctgatttt tttctattct acagccattt gatatgctgc aatcttaaca tatcaagtct 3600
cacgttggtg gacacaacat actatcacaa gtaagacacg aagtaaaacc aaccggcaac 3660

<210> 95

<211>

<212> DNA

<213> soy

15 ATGGATTCACTGCTTCTTCGATCTTTCCCTAATATTAATAACGCCTCTTCTCTCACCACCACTGGTGCAAATTTCTCC
AGGACTAAATCTTTGCGCAACATTTACCATGCAAGTTCTTATGTGCCAAATGCTTCATGGCACAATAGGAAAATCCAA
AAAGAATATAATTTTTTGGAGGTTTCGGTGGCCAAGTTTGAACCATCATTACAAAGGCATTGAGGGAGCGTGTACATGT
AAAAATGTAATATAAAATTTGTTGTGAAAGCGACCTCTGAAAAATCTCTTGAGTCTGAACCTCAAGCTTTTGATCCA
AAAAGCATTTTGGACTCTGTCAAGAATTCCTTGGATGCTTTCTACAGGTTTTCCAGGCCTCACACAGTTATTGGCACA
GCATTAAGCATAATTTCTGTGTCTCTTCTTGCTGTTGAGAAAATATCAGATATATCTCCATTATTTTTTACTGGTGTG
20 TTGGAGGCTGTGGTTGCTGCCCTGTTTATGAATATTTATATTGTTGGTTTGAATCAATTGTCTGATGTTGAAATAGAC
AAGATAAAACAAGCCGTATCTTCCATTAGCATCTGGGGAATATTCCTTTGAAACTGGTGTCACTATTGTTGCATCTTTT
TCAATTCTGAGTTTTTGGCTTGGCTGGGTTGTAGGTTTCATGGCCATTATTTTGGGCCCTTTTTGTAAGCTTTGTGCTA
GGAAGTCTTATTCAATCAATGTGCCTCTGTTGAGATGGAAGAGGTTTGCAGTGCTTGCAGCGATGTGCATTCTAGCT
GTTTCGGGCAGTAATAGTTCAACTTGCATTTTTTCCCTTCACATGCAGACTCATGTGTACAAGAGGCCACCTGTCTTTTCA
25 AGACCATTGATTTTTGCTACTGCATTCATGAGCTTCTTCTCTGTAGTTATAGCACTGTTTAAGGATATACCTGACATT
GAAGGAGATAAAGTATTTGGCATCCAATCTTTTTTCAGTGCGTTTAGGTGAGAAGCCGGTGTCTGGACTTGTGTTACC
CTTCTTGAAATAGCTTATGGAGTCGCCCTCCTGGTGGGAGCTGCATCTCCTTGTCTTTGGAGCAAATTTTACAGGGT
CTGGGACACGCTGTGCTGGCTTCAATTCTCTGGTTTCATGCCAAATCTGTAGATTGAAAAGCAAAGCTTCGATAACA
30 TCCTTCTATATGTTTATTTGGAAGCTATTTTATGCAGAATACTTACTCATTCTTTTGTAGATGA

<210> 96

<211>

<212> DNA

<213> soy

40 ATGGATTTCGATGCTTCTTCGATCTTTTCCTAATATTAACAACGCTTCTTCTCTCGCCACCACTGGTTCTTATTTGCCA
AATGCTTCATGGCACAATAGGAAAATCCAAAAGAATATAATTTTTTGGAGGTTTCGGTGGCCAAGTTTGAACCACCAT
TACAAAAGCATTGAAGGAGGGTGTACATGTAAAAATGTAATATAAAATTTGTTGTGAAAGCGACCTCTGAAAAATCT
TTTGAGTCTGAACCCCAAGCTTTTGATCCAAAAGCATTTTGGACTCTGTCAAGAATTCCTTGGATGCTTTCTACAGG
TTTTCCAGACCTCACACAGTTATTGGCACAGCATTAAAGCATAATTTCTGTGTCCCTCCTTGCTGTTGAGAAAATATCA
GATATATCTCCATTATTTTTTACTGGTGTGTTGGAGGCTGTGGTTGCTGCCCTGTTTATGAATATTTATATTGTTGGT
TTGAATCAATTGCTGATGTTGAAATAGACAAGATAAACAAGCCGTATCTTCCATTAGCATCTGGGGAATATTCCTTT
45 GAAACTGGTGTCACTATTGTTGCATCTTTTTCAATTCTGAGTTTTTGGCTTGGCTGGGTTGATAGTTTCATGGCCATTA
TTTTGGGCCCTTTTTGTAAGCTTTGTGCTAGGAAGTCTTATTCAATCAATGTGCCTCTGTTGAGATGGAAGAGGTTT
GCAGTGCTTGCAGCGATGTGCATTCTAGCTGTTTCGGGCAGTAATAGTTCAACTTGCATTTTTTCCCTTCACATCCAGACT
CATGTATACAAGAGGCCACCTGTCTTTTCAAGATCATTGATTTTTGCTACTGCATTCATGAGCTTCTTCTGTAGTT
ATAGCACTGTTTAAGGATATACCTGACATTGAAGGAGATAAAGTATTTGGCATCCAATCTTTTTTCAGTGCGTTTAGGT
CAGAAGCCGGTATTCTGGACTTGTGTTATCCTTCTTGAAATAGCTTATGGAGTCGCCCTCCTGGTGGGAGCTGCATCT
50 CCTTGTCTTTGGAGCAAATTTGTCACGGGTCTGGGACACGCTGTTCTGGCTTCAATTCTCTGGTTTCATGCCAAATCT
GTAGATTTGAAAAGCAAAGCTTCGATAACATCCTTCTATATGTTTATTTGGAAGCTATTTTATGCAGAATACTTACTC
ATTCCTTTTGTAGATGA

<210> 97

<211>

<212> PRT

5 <213> soy

MDSMLLRSPFNINNASSLATGSYLPNASWHNRKIQKEYNFLRFRWPSLNHHYKSIEGGCTCKKCNKIFVVKATSEKS
FESEPQAFDPKSILDSVKNSLDAFYRFSRPHTVIGTALSIIISVSLLAWEKISDISPLFFTGVLEAVVAALFMNIYIVG
10 LNQLSDVEIDKINKPYLPLASGEYSFETGVTIVASFILSFWLGVVGSWPLFWALFVSFVLGTAYSINVPLLRWKRF
AVLAAMCILAVRAVIVQLAFFLHIQTHVYKRPPVFSRSLIFATAFMSFFSVVIALFKDIPDIEGDKVFGIQSFSVRLG
QKPVFWTCVILLEIAYGVALLVGAASPCLSKIIVTGLGHAVLASILWFHAKSVDLKSKASITSFYMFIVKLFYAEYLL
IPFVR

<210> 98

15 <211>

<212> PRT

<213> soy

MDSLLLRSPFNINNASSLTTTGANFSRTKSFANIYHASSYVPNASWHNRKIQKEYNFLRFRWPSLNHHYKGIEGACTC
KKNKIFVVKATSEKSLESEPQAFDPKSILDSVKNSLDAFYRFSRPHTVIGTALSIIISVSLLAWEKISDISPLFFTGV
LEAVVAALFMNIYIVGLNQLSDVEIDKINKPYLPLASGEYSFETGVTIVASFILSFWLGVVGSWPLFWALFVSFVL
GTAYSINVPLLRWKRFVLAAMCILAVRAVIVQLAFFLHMQTHVYKRPPVFSRPLIFATAFMSFFSVVIALFKDIPDI
EGDKVFGIQSFSVRLGQKPVFWTCVILLEIAYGVALLVGAASPCLSKIIFTGLGHAVLASILWFHAKSVDLKSKASIT
5 SFYMFIVKLFYAEYLLIPFVR

<210> 99

<211>

<212> DNA

<213> rice

GAGCAGCACTGGGTCTTACATTCCAATGGAGCTCGCCTGTTGCTTTTCATTACATGCTTCG
TGACTTTTATTTGCTTTGGTCATTGCTATAACCAAAGATCTCCCAGATGTTGAAGGGGATC
GGAAGTATCAAATATCAACTTTGGCGACAAAGCTCGGTGTGAGAAACATTGCATTTCTTG
GCTCTGGTTTATTGATAGCAAATTATGTTGCTGCTATTGCTGTAGCTTTTCTCATGCCTC
35 AGGCTTTTCAGGCGCACTGTAATGGTGCCTGTGCATGCTGCCCTTGCCGTTGGTATAATTT
TCCAGACATGGGTTCTGGAGCAAGCAAAATATACTAAGGATGCTATTTACAGTACTACC
GGTTCATTTGGAATCTCTTCTATGCTGAATACATCTTCTTCCCGTTGATATAGAGACCAA
GCAATCTGATATGGTCTGCATGTTGAGTGCAGCAAAACTAGAAGCCCATATGAACAGTG
GGAGTAGGGGAACGAACATGCCATCCATGGGAAGACTCTGATAACTCTCTCTCGCCCGGG
40 CTGTAAGGGTAAGCACTGTTGGGCATATATATGAAAGGAAGGTGATAAAGCAGGGATGC
TAAATTGCTACTGGGATCCTCAAAGGCTTATAGTGGTCACCACTGGAATGTGCCTTAATA
ATTTGGTTACCCAGCAGAGCAAGTTTTTGAGGTTATTAGGTAATATCTTTGAGGGAATG
AACTTAGATTTTATTGTTTTAAGGTCTGGTCACACAACGGGTAGTAGTGTGGAGCGGCA
45 AAAACGACCTTGTTTTACACTACCAAGGGAGGTTAACTCTAGTTTTTATGTGACCACTT
ACCTTGAGAGTTGAGACCATGGAATCACTTGTGCACTCCTCGGCTGTATATTTCTAGTG
TCAGCATTTGCATTCTCTCCCACTTGTACTTGAAAAGTTGAAGACAACTTTTTGTGT
GTGT

<210> 100

50 <211>

<212> DNA

<213> wheat

CGTCCGCGGACGCGTGGGTGCTTATTCAGTCAATCTGCCGCACTTTCTATGGAAGAGATC
55 TGCTGTTGTTGCAGCACTCTGCATATTAGCAGTGGTGGGTGATAGTTCAACTGGCATT

TTTTCTCCACATTCAGACATTTGTTTTTCAGAAGGCCGGCAGACTTTTCAAAGCCATTGAT
ATTTGCAACTGCCTTCATGACATTTCTCTCAGTTGTAATAGCATTATTCAGGATATACC
CGATATTGAAGGGGACCGCATCTTTGGAATCCAATCTTTTAGTGGTAGACTAGGTCAAAG
CAGGGGTTTTCTGGACTTGCGTTGGCCTACTTGAGGTTGCCTACGGTGTGCGATACTGAG
5 GGGGTAACCTCTCCAGTTTGTGGAGCAAATCTATAACTGTTGTGGGCCATGCAATCCT
C

<210> 101

<211>

<212> DNA

<213> leek

GTTTCCCCCCTCGAATTTTTTTTTTTTTTTTTTACTTCATTTTTCTGTGAATAAATTCT
15 TAAAAAAGACAAAGAAAACCACTGGATATCCTAAATTCACATAGGCTATTGTCTATTCAA
TGATAATCTTTAACACAACATACAACATGAATATAATTAAGGAGAAATGATCTGCAATTG
TTGAAAGAAGCTCTCCGTTTTTAAGATGACAATTAAAGCGTTGTTAATTCCAGCCATTTCT
GCCTCCATTATCTACTCATCTTCTCTTGCGATTCTTTTCCATGTAGGTCTATAAACCTCA
TCTTACAAAAGGAATGAGCAAGTACTCAGCATAGAAGAGCTTCCACACGAACATATAAAA
20 AGATGTAATAGTGGTTTTTGGTCATTGGTCCATGAGATCTAGCACGATTCCAAAGTAACGA
CCCAAGAATTGCATGACCTATCACTGTTAAGCATTGCTCCATAGGCATGAGGAAGTAGC
TCCAACAACCATGACAACAGTGTAGGCCATCTCAAGGAGATATATACATATCCAAAACAC
CCTCTCCTGGCCAAGGCGCACGCTGAAAGAATGGATGCCAAATATTTTGTCTCCGTCTAT
ATCAGGTATATCCTTAAATAGAGCAATAACAACCTGAGAAGAAGCTCATGAAGGCAGTTGC
25 AAATATCAATGGCCTTGTGAAACTTGCTGGTCTTTTGAAAACAAA

<210> 102

<211>

<212> DNA

<213> leek

NATTCGGCAGCAGTTTTGAAGAAGTTAAGCATGGACTCCCTCCTTACCAAGCCAGTTGTA
ATACCTCTGCCTTCTCCAGTTTGTTCCTACTACCAATCTTGCGAGGCAGTTCTGCACCAGGG
CAGTATTCATGTAGAACTACAATCCAATAAGAATTCAAAGGTGCCTCGTAAATTATGAA
CATGTGAAACCAAGGTTTACAACATGTAGTAGGTCTCAAAAACCTGGTTCATGTAAAAGCC
ACATCCGAGCATTCTTTAGAATCTGGATCCGAAGGATACACTCCTAGAAGCATATGGGAA
GCCGTACTAGCTTCACTGAATGTTCTATACAAATTTTACGACCTCACACAATAATAGGA
ACAGCAATGGGCATAATGTCAGTTTCTTTGCTTGTGTCGAGAGCCTATCCGATATTTCT
CCTCTGTTTTTTGTGGGATTATTAGAGGCTGTGGTTGCTGCATTGTTTATGAATGTTTAC
40 ATTTAGGTCTGAATCAATTATTTGACATAGAAATAGACAAGGTCAATAAACCTGATCTT
CCTCTTGATCTGGAGAATACTCACCAAGAGCTGGTACTGCTATTGTCATTGCTTCAGCC
ATCATGAGCTTTGGCATTGGATGGTTAGTTGGCTCTT

<210> 103

<211>

<212> DNA

<213> canola

TTTTTTTTTTTTTTTTTCAAAAAGACCAATCCTTTAGTATGTACAT
GAACAAAGTGATTTTGTCTCCAAGCTACAAAGAAGAAGAGAGGTATACAAAGAAAAC
TACAAATGTTTACCATGAATGCTAGAAGAAGGGAATAACAGATACTCTGCCTAGAAGAG
ATTCCATATAAACCGTAATATCCTGCTATAGCTTCCTTTGTGTAGTTTGCTTTTTCTAG
CACCCATGTCTGGAACCAAGCATGAAGCCAAGATCATATGTGCAGGAATCATCAAGCT
ACCTCTAAAAACCTGAGGCATGTAGAAAGCTAGTGATATGGCAGAAATATAGTTCACTAG
55 CAGAAGTCCAGAACCAGGAATGCAATGTTCCCTCACTCCAAGCTTTGTTGCTAGTGTGA
TATTTGGAAGTTGCGATCTCCTTCAACATCAGGAAGATCTTTTGTAAATAGCAATGACTAG
TGCAACAGTGTACAAAAGACGTGATGAAAGCCACAGGTGCACTCCACTGAAACGAAAG

TCCAAGAGCAGCTCTAGTAGCATGGTACACACCAAAATTAAGAAGAAAACCTCGTACCGT
GGCAATAATAAGAAACGCTGCAACTGGAATCTCTTCAATCTAAATGGTGGAACAGAATA
GATGGTCCCCAGATCGGACGCGTGGGTCGAC

5 <210> 104
<211>
<212> DNA
<213> corn

10 CCACGCGTCCGCCCGGCCAAGGGATGGACGCGCTTCGCCTACGGCCGTCCTCCTCCCCG
TGCGGCCCGCGCGGCCCGGCCGCGAGATCATTTTCTACCACCATGTTGTTCCATACAAC
GAAATGGTGAAGGACGAATTTGCTTTTCTAGCCAAAGGACCAAGGTCTACCTTGCATC
ACCATCAGAAATCTTTCGAATGGAAATCCTCCTATTGTAGGATATCACATCGGTCATTAA
15 ATACTTCTGTTAATGCTTCGGGGCAACAGCTGCAGTCTGAACCTGAAACACATGATTCTA
CAACCATCTGGAGGGCAATATCATCTTCTCTAGATGCATTTTACAGATTTTCCCGGCCAC
ATACTGTATAGGAACAGCATTAAGCATAGTCTCAGTTTCCCTTCTAGCTGTCCAGAGCT
TGTCTGATATATCACCTTTGTTCTCTACTGGTTTGTCTGGAGGCAGTGGTAGCTGCCCTTT
TCATGAATATCTATATTGTTGGACTGAACCAGTTATTGACATTGAGATAGACAAGGTTA
20 ACAAGCCAACCTCTCCATTGGCATCTGGGGAATACACCCTTGCAACTGGGGTTGCAATAG
TTTCGGTCTTTGCCGCTATGAGCTTTGGCCTTGGATGGGCTGTTGGATCACAACTCTGT
TTTGGGCTCTTTTCATAAGCTTTGTTCTTGGGACTGCATATTCAATCAATCTGCCGTACC
TTGATGGAAGAGATTTGCTGTTGTTGCAGCACTGTGCATATTAGCAGTTTCGTGCAGTGA
TTGTTTTCAGCTGGCCTTTTTTCTCCACATTCAGACTTTTGTTCAGGAGACCGGCAGTGT
25 TTTCTAGGCCATTATTATTTGCAACTGGATTATGACGTTCTTCTCTGTTGTAATAGCAC
TATTCAAGGATATACCTGACATCGAAGGGACCGCATATTCGGGATCCGATCCTTCAGCG
TCCGGTTAGGGCAAAAGAAGGTCTTTTGGATCTGCGTTGGCTTGCTTGAGATGGCCTACA
GCGTTGCGATACTGATGGGAGCTACCTCTTCTGTTTGTGGAGCAAAACAGCAACCATCG
CTGGCCATTCCATACTTGCCGCGATCCTATGGAGCTGCGCGCGATCGGTGGACTTGACGA
30 GCAAAGCCGCAATAACGTCTTCTACATGTTTCTGGAAGCTGTTCTACGCGGAGTACC
TGCTCATCCCTCTGGTGGGTCGCGGAGCGGAGGCTGGTGGCAGACGGATCGGCGTCGG
CGGGGCGGCAAAACAACTCCACGGGAGAACTTGAGTGCCGGAAGTAACTCCCGTTGAAA
GTTGAAGCGTGCACCACCGGCACCGGGCAGAGAGACACGGTGGCTGGATGGATACGGA
TGCCCCCCCCAATAAATTCCTCCGTCATGGTAAAAAAAAAAAAAAAAAAAA

5 <210> 105
<211>
<212> DNA
<213> corn

40 GCCGCGCAGCGCGACGAGCGCCACCTGCTTGCTGCCGCGTGCTGCGTGCGTGTGCGTCC
ACCACTGACCCCGCGCCCGCCCGCCGCTGCCCTCCACTCCACTTGTCTACTCGTCCG
CGGCCCGCTTCCCCCCCCGCCAAGGGATGGACGCGCTTCGCCTACGGCCGTCCTCCTCC
CCGTGCGGCCCGCGCGGCCCGCCGCGAGATCATTTTCTACCACCATGTTGTTCCATAC
45 AACGAAATGGTGAAGGACGAATTTGCTTTTCTAGCCAAAGGACCAAGGTCTACCTTGC
ATCACCATCAGAAATCTTTCGAATGGAAATCCTCCTATTGTAGGATATCACATCGGTCT
TAAATACTTCTGTTAATGCTTCGGGGCAACAGCTGCAGTCTGAACCTGAAACACATGATT
CTACAACCATCTGGAGGGCAATATCATCTTCTCTAGATGCATTTTACAGATTTTCCCGGC
CACATACTGTATAGGAACAGCATTAAGCATAGTCTCAGTTTCCCTTCTAGCTGTCCAGA
50 GCTTGTCTGATATATACCTTTGTTTCTCTACTGGTTTGCTGGAGGCAGTGGTAGCTGCC
TTTTCATGAATATCTATATTGTTGGACTGAACCAGTTATTGACATTGAGATAGACAAGG
TTAACAAGCCAACCTCTCCATTGGCATCTGGGGAATACACCCTTGCAACTGGGGTTGCAA
TAGTTTTCGTCTTTGGCGCTATGAGCTTTGGCCTTGGATGGGCTGTTGGATCACAACTC
TGTTTTGGGCTCTTTTCATAAGCTTTGTTCTTGGGACTGCATATTCAATCAATCTGCCGT
ACCTTCGATGGAAGAGATTTGCTGTTGTTGCAGCACTGTGCATATTAGCAGTTTCGTGCAG
55 TGATTGTTTCAGCTGGCCTTTTTTCTCCACATTCAGACTTTTGTTCAGGAGACCGGCAG
TGTTTTCTAGGCCATTATTATTTGCAACTGGATTATGACGTTCTTCTCTGTTGTAATAG
CACTATTCAAGGATATACCTGATCGAAGGGACCGCATATTCGGGATCCGATCCTTCA
GCGTCCGGTTAGGGCAAAAGAAGGTCTTTTGGATCTGCGTTGGCTTGCTTGAGATGGCCT
60 ACAGCGTTGCGATACTGATGGGAGCTACCTCTTCTGTTTGTGGAGCAAAACAGCAACCA
TCGCTGGCCATTCCATACTTGCCGCGATCCTATGGAGCTGCGCGCGATCGGTGGACTTGA
CGAGCAAAGCCGCAATAACGTCTTCTACATGTTTCTGGAAGCTGTTCTACGCGGAGT

ACCTGCTCATCCCTCTGGTGCGGTGAGCGCGAGGCGAGGTGGTGGCAGACGGATCGGCGT
CGGCGGGGCGGCAAACTCCACGGGAGAACTTGAGTGCCGGAAGTAACTCCCCTTTG
AAAGTTGAAGCGTGCACCACCGGCACCGGCAGAGAGACACGGTGGCTGGATGGATAC
GGATGGCCCCCAATAAATCCCCCGTGCATGGTACCCACGCTGCTTGATGATATCCC
5 ATGTGTCCGGGTGACCGGACCTGATCGTCTCTAGAGAGATTGGTTGCACAACGTCCAACA
TAGCCCGTAGGTATTGCTACCACTGCTAGTATGATACTCCTTCCTAGTCCTTGCCAGCAC
CAGTGACCCAACTTGGTCCGCTGAGCTCAGCGCTCAGCAGCTTTACGTGCATCTGCGCC
TTGACTTGTGCAGTGGGCGTCGCTAGCATGAATGATGTATGGTGCCTCAGGCGCTGACGG
10 TTCGTGAGTCTGGGCGGTGTTTGTGTCCGAGGAAGATCGTCTGTGAGAGATCTGGATTG
CCTCGCTGCT

<210> 106

<211>

<212> DNA

<213> corn

CGGCCGGACTCTTCTGACTTGGCAACCGCGCGCAGCGGACGAGCGCCACCTGCTTGCT
GCCGCGTGCCGTGCGTGTGCGTCCACCACTGACCCCGCGCCCGCCGCGCCCTGC
20 CCTCCACTCCACTTGCTCACTCGTCGGCTCGTCGCGGCCCGCTCCCCCGGCCAAGG
GATGGACGCGCTTCGCCCTACGGCCGTCCCTCCTCCCGTGCGGCCCGCGCGGCCCGCCC
GCGAGGCAGTGGTAGCTGCCCTTTTCATGAATATCTATATTGTTGGACTGAACCAGTTAT
TCGACATTGAGATAGACAAGGTTAACAAGCCAACCTCTCCATTGGCATCTGGGGAATACA
CCCTTGCAACTGGGGTTGCAATAGTTTCGGTCTTTGCCGCTATGAGCTTTGGCCTTGGAT
5 GGGCTGTTGGATCACAACCTCTGTTTTGGGCTCTTTTCATAAGCTTTGTTCTTGGGACTG
CATATTCAATCAATCTGCCGTACCTTCGATGGAAGAGATTGCTGTGTTGCAGCACTGT
GCATATTAGCAGTTCGTGCAGTGATTGTTTCAGCTGGCCTTTTTTCTCCACATTGAGACTT
TTGTTTTTCAGGAGACCGGCAGTGTTTTCTAGGCCATTATTAT

<210> 107

<211>

<212> DNA

<213> cotton

CCCACGCGTCCGAACATTGTTTGCACTTGTTATTGCCATAACCAAGGATCTTCCAGATGT
AGAAGGAGATCGCAAATTTCAAATATCAACATTAGCAACAAAGCTTGGAGTTAGAAATAT
TGCATTTCTTGGTTCCGGACTTCTACTGCTGAATTATGTTGCTGCTGTGTTGGCTGCAAT
ATACATGCCTCAGGCTTTCAGGCGTAGTTTAATGATACCTGCTCATATCTTTTGGCGGT
5 CTGCTTGATTTTTTCAGACATGGGTGTTGGAACAAGCAAATTACAAAAAGGAAGCAATCTC
GGGGTTCTATCGTTTTCATATGGAATCTCTTCTATGCAGAGTATGCGATTTTCCCCTTCGT
40 GT

<210> 108

<211>

<212> DNA

<213> tomato

CAGATCAATTCCAGTTCCTGCTGAGTTTTCTCCACTCAAACAGTTACATGCAATAGT
ACGGGTTTTGAAATGTAAAGCATGGAAGAGACCAAAAAAGCACTATTCCTCTCAATGAA
5 GTTGACAGCGGCAGTATATCACGCAAGAGCATGTTGGAGGAAGTGATCTAAGCACTATTGC
TGCTGATAAAAACTTAAAGGGAGATTTTTGGTGCACGCATCATCTGAACACCCTCTTGA
ATCTCAACCTTCTAAAGTCCTTGGGACTCAGTTAATGATGCCGTAGATGCTTTCTACAG
GTTCTCGCGGCCCATACCATAATAGGAACAGCATTGAGCATAAATTCAGTTCTCTCCT
TGCAGTTGAGAAGTTCTCTGATTTTTCTCCATTATTTTTCACTGGGGTGTTAGAGGCCAT
10 TGTTGCTGCCCTATTCATGAACATTACATAGTTGGTTTAAACCAGTTGTTCTGACATCGA
AATAGACAAGGTAAACAAGCCATATCTCCATTGGCATCAGGGGAATACTCTGTACAAAC
55 TGGAGTGATTGTTGTGTCGTCTTTTGCCATTTTGA

<210> 109

<211>

<212> DNA

<213> *Arabidopsis*

5 AACACCAAACACACAATTTACATTCTTTTGCATATTTCTTCTTCTTCTTCCATTATGGA
GATACGGAGCTTGATTGTTTCTATGAACCCTAATTTATCTTCCTTTGAGCTCTCTCGCCC
TGTATCTCCTCTCACTCGCTCACTAGTTCCGTTCCGATCGACTAAACTAGTTCCCCGCTC
CATTCTAGGGGGATCCCGTCGATCTCCACCCCGAATAGTGAAACTGACAAGATCTCCGT
TAAACCTGTTTACGTCCCGACGTCTCCCAATCGCGAACTCCGGACTCCTCACAGTGGATA
10 CCATTTTCGATGGAACACCTCGGAAGTTCTTCGAGGGATGGTGGATCCGGGTTTCCATCCC
AGAGAAGAGGGAGAGTTTTTGTGTTTATGTATTCTGTGGAGAATCCTGCATTTTCGGCAGAG
TTTGTCAACCATTTGGAAGTGGCTCTATATGGACCTAGATTCACTGGTGTGAGCTCAGAT
TCTTGGCGCTAATGATAAATATTTATGCCAATACGAACAAGACTCTCACAATTTCTGGGG
AGATCGACATGAGCTAGTTTTGGGGAATACTTTTAGTGCTGTGCCAGGCGCAAAGGCTCC
15 AAACAAGGAGGTTCCACCAGAGGAATTTAACAAGAAGAGTGTCCGAAGGGTTCCAAGCTAC
TCCATTTTGGCATCAAGGTCACATTTGCGATGATGGCCGTACTGACTATGCGGAAACTGT
GAAATCTGCTCGTTGGGAGTATAGTACTCGTCCCGTTTACGGTTGGGGTGATGTTGGGGC
CAAACAGAAGTCAACTGCAGGCTGGCCTGCAGCTTTTCCTGTATTTGAGCCTCATTGGCA
GATATGCATGGCAGGAGGCCTTTCCACAGGGTGGATAGAATGGGGCGGTGAAAGGTTTGA
20 GTTTCGGGATGCACCTTCTTATTCAGAGAAGAATTGGGGTGGAGGCTTCCAAGAAAATG
GTTTTGGGTCCAGTGTAATGTCTTTGAAGGGGCAACTGGAGAAGTTGCTTTAACCGCAGG
TGGCGGGTTGAGGCAATTGCCTGGATTGACTGAGACCTATGAAAATGCTGCACTGGTTTG
TGTAACACTATGATGGAAAAATGTACGAGTTTGTTCCTTGAATGGTGTGTTAGATGGGA
AATGTCTCCCTGGGGTTATTGGTATATACTGCAGAGAACGAAAACCATGTGGTGGAACT
25 AGAGGCAAGAACAATGAAGCGGGTACACCTCTGCGTGCTCCTACCACAGAAGTTGGGCT
AGCTACGGCTTGCAGAGATAGTTGTTACGGTGAATTGAAGTTGCAGATATGGGAACGGCT
ATATGATGGAAGTAAAGGCAAGGTGATATTAGAGACAAAGAGCTCAATGGCAGCAGTGG
GATAGGAGGAGGACCGTGGTTTGGGACATGGAAAGGAGATACGAGCAACACGCCCCGAGCT
ACTAAACAGGCTCTTCAGGTCCCATTTGGATCTTGAAAGCGCCTTAGGTTTGGTCCCTTT
30 CTTCAAGCCACCGGGTCTGTAACATTGATGAGTGTTTTGTTTGTGATAGAGACCCATGT
GATGAATGAAGCCTTAGTCATGTCATTGCTAGCTTCACTATTATGTATGTATGATTTTAG
TTCGTTTCGGTCCCTTGTGGTAAATGATACGGGCCAGTGTAAGT

<210> 110

35 <211>

<212> PRT

<213> *Arabidopsis*

40 MEIRSLIVSMNPNLSSFELSRPVSPLTRSLVPFRSTKLVPRSISRVSASI
STPNSETDKISVKPVYVPTSPNRELRTPHSGYHFDGTPRKFFEGWYFRVS
IPEKRESFCFMYSVENPAFRQSLSPLEVALYGPRFTGVGAQILGANDKYL
CQYEQDSHNFWDGRHELVLGNTFSAPVPAKAPNKEVPPEEFNRRVSEGFQ
ATPFWHQGHICDDGRDYAETVKSARWEYSTRPVYGWGDVGAKQKSTAGW
PAAFPVFEPHWQICMAGGLSTGWIEWGGERFEFRDAPSYSEKNWGGGFPR
45 KWFVWQCNVFEGATGEVALTAGGGLRQLPGLTETYENAALVCVHYDGKMY
EFVPWNGVVRWEMSPWGYWYITAENENHVVELEARTNEAGTPLRAPTTEV
GLATACRDCSYGELKLQIWERLYDGSKGKVILETKSSMAAVEIGGGPWFG
TWKGDTSNTPELLKQALQVPLDLESALGLVPFFKPPGL*